Evolution and adaptation of Pseudomonas aeruginosa in cystic fibrosis airways

cystic fibrosis as a model system

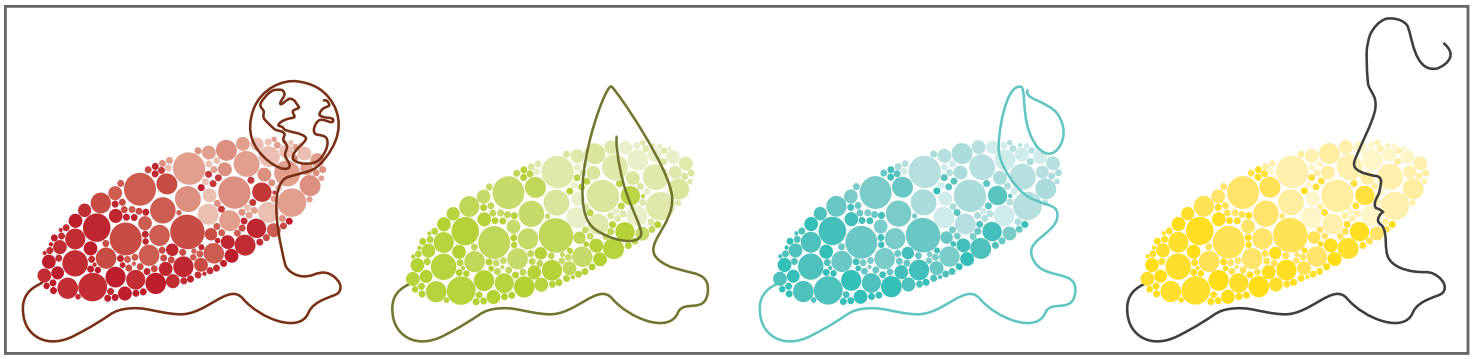
Madsen Sommer, Lea Mette

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Evolution and adaptation of *Pseudomonas aeruginosa* in cystic fibrosis airways

CYSTIC FIBROSIS AS A MODEL SYSTEM



Lea Mette Madsen Sommer
PhD Thesis

Novo Nordisk Foundation Center for Biosustainability & Department of Systems Biology
Technical University of Denmark, September 2015

Lea Mette Madsen Sommer, *Evolution and adaptation of Pseudomonas aeruginosa in cystic fibrosis airways, –Cystic fibrosis as a model system*. PhD Thesis.

PRIMARY SUPERVISOR:
Søren Molin, Professor

CO-SUPERVISORS:
Helle Krogh Johansen, DMSc
Lars Jelsbak, Associate Professor

LOCATION:
2970 Hørsholm

TIME FRAME:
October 2012 - September 2015

There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

There is another theory which states that this has already happened.

– Douglas Adams (The Restaurant at the End of the Universe)

PREFACE

This thesis is submitted as a partial fulfilment of the requirenents to obtain a PhD degree at the Technical University of Demark (DTU). The work was carried out from October 2013 to September 2015 primarily in the Cystic Fibrosis "Børn" (edt. Children) Infection Group (CF-Binfect) at the Novo Nordisk Foundation Center for Biosustainability (CfB) and partially at the Infection Microbiology Group (IMG) of Associate Professor Lars Jelsbak (co-supervisor), at The Department of Systems Biology, DTU.

The work was carried out under the supervision of Professor Søren Molin (supervisor), and DMSc Helle Krogh Johansen (co-supervisor) and funded by a PhD stipend from DTU.

Lea Mette Madsen Sommer

Skovlunde, September 2015

ABSTRACT

For centuries evolution has been investigated with an "end-point" approach, through comparisons between species or fossil records. However, to understand processes in general, including evolution, it is highly valuable to observe the dynamics as they unfold, in "real-time".

This is possible through laboratory experiments, with a high degree of control and rigour. But to truly understand evolution and the complex mechanisms it deploys, it is necessary to combine the laboratory learnings with investigations of natural systems. –Though, this can be tricky. Because of the heterogeneity and constant change of natural environments, the primary obstacle is re-sampling of the same population over time, especially if the population is small.

Nevertheless, it has been accomplished: Chronic airway infections of cystic fibrosis (CF) patients have offered a unique view into the adaptation and evolution of *Pseudomonas aeruginosa* to this natural environment, spanning thousands of bacterial generations. Because of the prolonged and persistent infections, they provide a valuable model system for the investigation of evolutionary mechanisms.

The main focus of this thesis has been to show the link between evolutionary studies in the CF model system and general evolutionary theories, many of which have been developed from observations of other organisms.

This comparison has initially been sought by showing the plausibility of using comprehensive collections of longitudinally sampled single isolates, for their use in evolutionary studies (Study 1). This was done by comparing five metagenomes with single isolates from four CF patients, and identifying significant genetic links found within the patient specific *P. aeruginosa* populations. This evident genetic link was even found for two populations, where a recent patient-to-patient transmission had occurred.

Secondly a comprehensive collection of 474 longitudinal single *P. aeruginosa* isolates from 34 young Danish CF patients was investigated by whole genome sequencing (Study 2). This was done to reconstruct the recent evolutionary history, and identify genes targeted in the initial adaptation to the CF airways. From this analysis we found common clonal lineages among the patients, evidence of patient-to-

patient transmission, historic contingencies, and convergent evolution of 52 candidate pathoadaptive genes.

By further genome sequencing 26 *P. aeruginosa* isolates from four Italian CF patients (Study 3), and 35 *P. aeruginosa* isolates from 12 primary ciliary dyskinesia (PCD) patients (Study 4), we were able to find genetic and phenotypic links across countries and diseases. All three studies (not including the metagenome study) had common clonal lineages and clear overlaps of genetic adaptational patterns. However, the genetic overlap between CF and PCD isolates did not extend to a phenotypic overlap, which indicates that the mucus, which is different in CF patients compared to PCD patients, is a significant selective factor for the evolution and adaptation of *P. aeruginosa* to these environments.

Independently and together the studies presented in this thesis provide new knowledge of adaptation and evolution in both CF and PCD airways. With further characterisation of genetic and phenotypic adaptations it should be possible to translate these results into clinically relevant information, leading to better epidemiological predictions, valuable information with regards to treatment strategies, and perhaps extrapolation of this knowledge to other infection scenarios.

OVER ALL: Through the convergence of genetic and phenotypic adaptations observed in CF studies and by linking processes of evolution to these observations, this thesis shows that collections of longitudinal *P. aeruginosa* isolates from CF patients provide a valuable basis for the study of adaptation and evolution in natural environments.

DANSK RESUMÉ

I århundreder har man undersøgt evolution ved at bruge en "endpoint" fremgangsmåde. Det vil sige, at man primært har lavet sammenligninger imellem arter og ved lejlighed fossiler. Men for at kunne forstå processer generelt, inklusiv evolution, er det uvurderligt at kunne undersøge processen undervejs og derved observere den "live".

Dette er muligt gennem laboratorieeksperimenter på grund af den høje grad af kontrol og klarhed. Men for at få en sand forståelse af evolution og de komplekse mekanismer det indebærer, er det nødvendigt at kombinere laboratorieforsøg med undersøgelser af naturlige miljøer.

På grund af heterogeneciteten og den konstante forandring af naturlige miljøer er den primære forhindring i dette tilfælde observationen af den samme population over tid, specielt hvis populationen er lille.

Ikke desto mindre har det vist sig at være muligt: Kroniske luftvejsinfektioner hos cystisk fibrose (CF) patienter har vist sig at give et enestående indblik i adaptation og evolution af *Pseudomonas aeruginosa* til dette naturlige miljø. Det har tilmed været muligt at følge denne proces over tusinder af bakterielle generationer. På grund af den langvarige og persisterende kolonisering og infektion af disse patienter, er CF infektionssystemet en værdifuld model til undersøgelser af evolutionære processer.

Det primære fokus for denne afhandling har været at vise links imellem evolutionære studier fra CF modelsystemer og generelle evolutionsteorier, hvoraf de fleste oprindeligt blev udviklet via observationer fra andre organismer.

Denne sammenligning blev i første omgang forsøgt ved at påvise muligheden for at bruge omfattende longitudinelle isolat samlinger til studier af evolution (Studie 1). Dette blev opnået ved sammenligning af fem sputummetagenomer med longitudinelle enkeltisolater fra fire CF patienter og identifikationen af klare genetiske links indenfor den enkelte patients specifikke *P. aeruginosa* population. Dette genetiske link mellem enkelte isolater og metagenomer blev tilmed påvist for to populationer, hvor en umiddelbar patient-til-patient transmission havde foregået og derved øget ligheden imellem de to populationer signifikant, i forhold til populationerne fra de andre patienter.

Efterfølgende blev en omfattende samling af 474 longitudinelle *P. aeruginosa* isolater udsat for hel-genom sekvensering (Studie 2). Disse isolater stammede fra 34 unge Danske CF patienter. Denne samling blev analyseret for at rekonstruere den umiddelbare evolutionære historie og identificere gener, der var vigtige for den indledende adaptation til CF luftveje. Fra denne analyse identificerede vi fælles klontyper mellem patienter, beviser for patient-til-patient transmission, historiske sammenhænge mellem mutationer og konvergerende evolution af 52 kandidat pathoadaptive gener.

Yderligere sekvensering, af 26 *P. aeruginosa* isolater fra fire Italienske CF patienter (Studie 3) og 35 *P. aeruginosa* isolater fra 12 primær cilie dyskinesi (PCD) patienter (Studie 4), gjorde os i stand til at finde genetiske og fænotypiske links på tværs af landegrænser og sygdomme. Alle tre studier (ikke inkluderende metagenom studiet) havde fælles klontyper og klare overlap af adaptive mutationer. Men det genetiske overlap imellem CF og PCD isolater kunne ikke udvides til at inkludere fænotypiske overlapninger, hvilket indikerer, at det slim, der findes i luftvejene, og som er markant anderledes for CF patienter, er en signifikant selektionsfaktor for evolutionen og adaptationen af *P. aeruginosa* i disse miljøer.

Uafhængigt af hinanden, men også i kombination giver de studier præsenteret i denne afhandling ny og udvidet viden omkring adaptation og evolution til både CF og PCD luftveje. Ved yderligere karakterisering af genetiske og fænotypiske adaptationer burde det være muligt at oversætte disse resultater til klinisk relevant information, der kan lede til vigtig viden med hensyn til behandlingsstrategier og muligvis også ekstrapolation til andre infektionsscenarier.

OVERORDNET: Igennem den konvergerende genetiske og fænotypiske adaptation, som er observeret i CF studier og ved at linke dette til evolutionære processer viser denne afhandling at samlinger af longitudinelle *P. aeruginosa* isolater fra CF patienter giver en værdifuld basis for studier af adaptation og evolution i naturlige miljøer.

RESEARCH ARTICLES

WORK INCLUDED IN THIS THESIS

Study 1:
Sommer L. M., Marvig R. L., Luján A., Koza A., Pressler T., Molin S., and Johansen H. K. (2015) Investigations of inter- and intra-clonal diversity of *Pseudomonas aeruginosa* populations in cystic fibrosis patients. *Manuscript in preparation.*

Study 2:
Marvig R. L., **Sommer L. M.**, Molin S., and Johansen H. K. (2015) Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nature Genetics*, **47**: p. 57-64. doi:10.1038/ng.3148

Study 3:
Marvig R. L.*, Dolce D.*, **Sommer L. M.***, Petersen B., Ciofu O., Campana S., Molin S., Taccetti G., and Johansen H. K. (2015) Within-host microevolution of *Pseudomonas aeruginosa* in Italian cystic fibrosis patients. *Manuscript submitted for publication.*

Study 4:
Sommer L. M.*, Alanin M. C.*, Marvig R. L., Nielsen K. G., Høiby N., von Buchwald C., Molin S., and Johansen H. K. (2015) Evolutionary and adaptational differences of *Pseudomonas aeruginosa* in primary ciliary dyskinesia and cystic fibrosis patients. *Manuscript in preparation.*

* Denotes equal contribution.

PUBLISHED WORK NOT INCLUDED IN THIS THESIS

Marvig R. L., **Sommer L. M.**, Jelsbak L., Molin S., and Johansen H. K. (2015) Evolutionary insight from whole-genome sequencing of *Pseudomonas aeruginosa* from cystic fibrosis patients *Future Microbiology*, **10**(4): p. 599-611. doi:10.2217/fmb.15.3

ACKNOWLEDGEMENTS

Acknowledgements are my least favourite part of writing, because I am not sure if I forget somebody, and how to write this without it getting to tacky. So, if this is the case I am sorry. Everyone I really need to thank, you know who you are and I am truly grateful for what you have done for me.

So to begin: A thank is definitively in place for my main supervisors Søren Molin and Helle K. Johansen. Thank you for guiding me through the endless roads my project could have, and has lead me down to a final destination that I am proud to have reached, and that I could not have reached without your help. This process would also not have been possible had it not been for the immense help and patience of Rasmus L. Marvig, who introduced me to, and has taught me all I know about bioinformatics (almost).

While we are at bioinformatics, then I will also thank Sandra W. Thrane for helping me reach a further understanding by discussing and sharing frustrations about this topic. Including frustrations about computer language and the general lack of information of this topic in research papers (come on guys!).

Furthermore, I also thank Trine M. Markussen, Alicia J. Fernandes, Sandra B. Andersen, Karen I. Starlit, Mette Munk, and Rasmus for a good office atmosphere, fantastic humour, and nice coffee break talks about all and nothing.

I will especially like to thank Eva K. Andresen and Linda R. Jensen, for their ability to get me cheered up, despite the continuous frustrations of science, and Eva also for being my biking buddy in all kinds of weather. Also, I thank Linda, Rasmus, and Sandra (both of you) for the discussions of new ideas, texts, and figures.

Not to forget collaborators outside of DTU: Thank you, Ulla Johansen and Nikolai Kirkby for your help at the University Hospital of Copenhagen during the purification of hundreds (must be close to a thousand!) *Pseudomonas aeruginosa* genomes. In this connection, again, thank to Helle for the great work of building the comprehensive Danish CF children isolate collection. Also from the University Hospital of Copenhagen, Mikkel C. Alanin, from The University Hospital in Florence, Daniela Dolce, and from Exeter University in England, Adela Luján.

Thank you for great collaborations.

Because of the move to Hørsholm and the NNF center in the middle of my PhD studies there are a lot of people with whom I have shared an office and workspace, and therefore thank for the general discussions and talks. This includes both past and present members of Lars Jelsbak's group IMG (at Lyngby campus) and CFBinfect (in Hørsholm). In particular: Susanne (Søs) Kofoed, Mikkel Lindegaard, Rikke K. Lauridsen, Fatima A. Atraktchi, Katherine Long, Juliane C. Thøgersen, and Anne-Mette Christiansen. Including the most recent contributions to the CFBinfect group: Janus A. Haagensen, Anne Loch, and Jennifer Bartell.

I also send my regards to Associate Professor Jeffrey Barrick and the people in his lab at The University of Texas, for their help during my 1.5 month stay in Austin, Texas (USA).

Last by not least, a huge thank is to be directed at my family and friends who have accepted my absentness during the process of writing this thesis as well as helping me with my ups and downs the last three years. In particular my husband Kenneth Sommer who has lived with my craziness and supported me, in particular, the last couple of months where the hard work of three years culminated in the writing of this thesis.

CONTENTS

LIST OF FIGURES

ABREVIATIONS

ASM	Artificial sputum medium
AES-2	Australian epidemic strain-2
CF	Cystic fibrosis
CFTR	Cystic fibrosis trans-membrane conductance regulator
dN	Rate of nonsynonymous genetic change
dS	Rate of synonymous genetic change
HGT	Horizontal gene transfer
LES	Liverpool epidemic strain
MIC	Minimum inhibitory concentration
MRCA	Most recent common ancestor
PCD	Primary ciliary dyskinesia
SNP	Single nucleotide polymorphism

1

OVERVIEW AND STRUCTURE OF THESIS

1.1 OVERVIEW

This thesis is meant to give a general description of cystic fibrosis (CF) as a model system, where continuous sampling of bacterial populations can be used for the investigations of adaptation and evolution to new environments.

Together with laboratory experiments, this model system makes it possible to identify the impacts of different evolutionary mechanisms and processes in a natural environment. This is important both in clinical settings and in industrial settings, for the understanding and development of improved production strains and treatment strategies.

1.2 STRUCTURE OF THESIS

This thesis is organised in four chapters:

Chapter 1 (the current chapter) is a general introduction to the thesis as a whole.

Chapter 2 includes a theoretical background to evolutionary processes and link these to previous findings from studies of CF, as well as the major findings of the papers included in this thesis. This is done by including short descriptions of the papers (Study 1 to 4), when necessary.

Chapter 3 includes the overall conclusions and future perspectives.

Chapter 4 encloses the full-length research articles included in this thesis.

2 CYSTIC FIBROSIS AS A MODEL SYSTEM

2.1 INTRODUCTION

Evolution is a process that has shaped the world and the remarkable diversity we see around us, and the ability to understand this process will take us closer to understanding life as we see it today. This understanding includes the bigger questions of how life has evolved, but also the smaller, more pressing questions, of resistance development in microorganisms and how to develop the best production strain for a given product, which could be an antibiotic.

For centuries evolution has been investigated by comparative studies of living organisms, by investigating differences between species originating from the same ancestor, including comparisons with fossil records. The common denominator being the "end-point" approach focusing on fundamental differences or similarities between species, not as they evolve but as they *have* evolved [1, 2]. However, to understand processes in general, including evolution, it is highly valuable to observe the dynamics as they unfold, on account of the degree of detail that cannot be found by "end-point" comparisons[3].

With their manageable genomic complexity and fast growth, bacteria make ideal model organisms and because of the advent of modern technologies we are now capable of observing and investigating evolution as it is unfolding in "real-time", both phenotypically and genotypically. This entails investigations of processes of speciation, where invasion of new niches bring about alternate selective pressures forcing the bacterial organisms to adapt and evolve through natural selection.

To explain, NATURAL SELECTION acts on variation within a population by increasing the abundance of organisms with the highest reproductive success, also called FITNESS, in a given environment of specific SELECTIVE PRESSURES. In other words, when an organism is subjected to a selective pressure (e.g. antibiotics) it needs to adapt (e.g. antibiotic resistance), and this is done through the force of natural selection[3, 4].

By the use of bacterial populations it is possible to observe the nature and mechanisms of interactions between an organism and its surroundings, which forms the basis of adaptation and evolution. How-

ever, in most natural environments, if not all, the complexity of evolution and the difficulties of re-sampling the same population over time obscure the patterns of interaction. But without the complexity, posed by natural environments, the interplay and the complete picture cannot be achieved[3].

There is one natural environment where the study of evolution in "real time" has been found possible: cystic fibrosis (CF) airway infections. By investigating airway infections of patients with CF it has been possible to follow the same clonal lineage over a time period of approximately 35 years[5].

However, the complexity of natural systems, such as the CF airway infections, results in difficulties of data interpretation. To abate this, it is necessary to combine these learnings with those gained from controlled and rigorous laboratory studies. Through this fusion of knowledge it is possible to deduce theories concerning the intricate interplays of evolutionary processes, important for the understanding of life as we see it today.

When investigating evolution through bacterial populations it is important to be aware that not all mechanisms can be directly translated to all other organisms. Primarily because the mechanisms of speciation, diversity maintenance, and diversity generation in organisms with sexual reproduction can be markedly different from asexual organisms, such as bacteria[6, 7, 8].

Nevertheless, in many cases this extrapolation is possible, which will be illustrated by connecting investigations of bacterial populations from CF infections, with general evolution theories, – which in most cases have been developed from observations of other organisms. Through the same connections the possibility of using CF infections as model systems for the investigation of evolutionary processes will also be evident.

2.2 LABORATORY EVOLUTION

Laboratory evolution studies, which posses a high degree of control and rigour, are one aspect of the investigation of evolution. This is achieved by the simple set-up, where a population, or multiple populations is established from a common ancestor, which can then be propagated either by serial transfers or in chemostats (Figure 1). Both the ancestor as well as subsequent serially collected samples are stored as time passes (in most cases at -80°C) as a "fossil record". The stored samples can then be utilised in additional experiments

of how evolution and adaptation has progressed, by phenotypic and genotypic analyses. For quantitative measurements of evolution, fitness can be assessed by competition assays between ancestors and evolved lines[3]. With fitness, in this case, being the ability to produce descendants or survive under a specific selective pressure, such as the presence of antibiotics. The most important aspect of fitness measurements is to measure it in the environment where the organisms were initially evolved[9, 3].

All in all, this active investigation gives a more detailed understanding of the process of evolution than passively comparing traits between extinct and present day organisms.

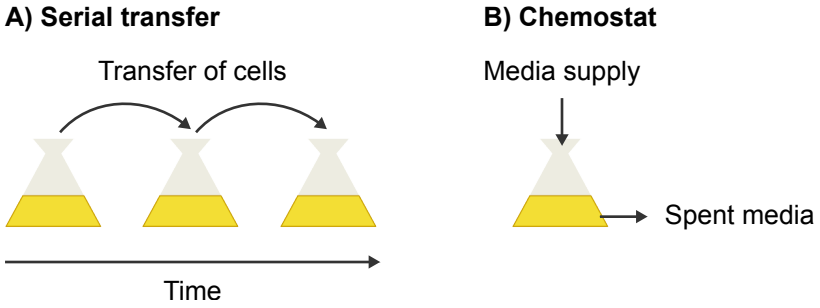


Figure 1: **Serial Transfer and Chemostats. A)** In a serial transfer experiment a bacterial isolate is cultured, and part of this culture is later transferred to fresh medium when appropriate. **B)** In a chemostat, fresh medium is continuously added to the system and liquid culture is continuously removed to keep the volume constant. It is possible to control the growth rate by changing the rate of media flowing into the system, giving additional control as compared to the serial transfer.

The longest running laboratory evolution experiment was set up in 1988 with *Escherichia coli* as the study organism and it has been running ever since, corresponding to more than 60,000 bacterial generations[3, 10]. This is carried out in a controlled environment of serial transfers of 12 lineages originating from the same ancestor.

This use of low-complexity laboratory environments surpass the problems found when investigating evolution in natural systems, which are much more complex.

However, in nature, we rarely find organisms as isolated pure cultures or in meticulously controlled environments, making the findings of laboratory experiments difficult to extrapolate. Therefore, to fully understand evolution it is necessary to pair the laboratory findings with findings from natural systems. Because only in natural and complex environments are the intricate interplays between the evolu-

tionary mechanisms fully present.

2.3 EVOLUTION IN NATURAL ENVIRONMENTS

Continuous sampling of a single or a few clonal lineages from natural populations is generally difficult, - if not impossible. Either because of temporal fluctuations and physical disturbances or because the population of interest is small.

Despite these difficulties of re-sampling, one natural system has proven to be useful: By investigating airway infections of patients with CF, it has been possible to re-sample the same bacterial lineage of *Pseudomonas aeruginosa*, also called clone type DK2, over a time period accounting for more than 200,000 bacterial generations[5].

CF is an autosomal recessive disease, which is caused by mutations in the Cystic Fibrosis Trans-membrane Conductance Regulator (CFTR) gene, encoding a Chloride-ion channel[11]. The mutation of the Cl^- channel causes it to malfunction to a higher or lower degree[12] resulting in a thick dehydrated mucus on all mucociliary surfaces in the body. Among other things, this leads to an impaired mucociliary clearance of the upper airways and thence an elevated risk of airway infections[13, 14]. In congruence with this, CF patients are continuously colonised and infected by a multitude of microbes, with the most prevalent being *Staphylococcus aureus* and *Haemophilus influenzae* in childhood and *P. aeruginosa* in adulthood (Figure 2)[15, 16].

Because of this high rate of bacterial infections CF patients are treated with large amounts of antibiotics[17], resulting in increased resistance development of the infecting species.

This thesis will primarily be focused on *P. aeruginosa*, since this is the organism found to cause colonisation and chronic infection in the majority of CF patients (Figure 2); It has a prevalence among the 26+ years old CF patients of >70% [15, 16], thus providing a broad basis for evolutionary studies. Furthermore it is able to cause colonisation and infection in a variety of other settings, such as chronic wounds and general nosocomial infections[18, 19]. Also, it has the ability to rapidly gain resistance towards multiple antibiotics[20], causing increased morbidity and mortality for infected patients[19]. Thereby making *P. aeruginosa* highly relevant for infection studies, as well.

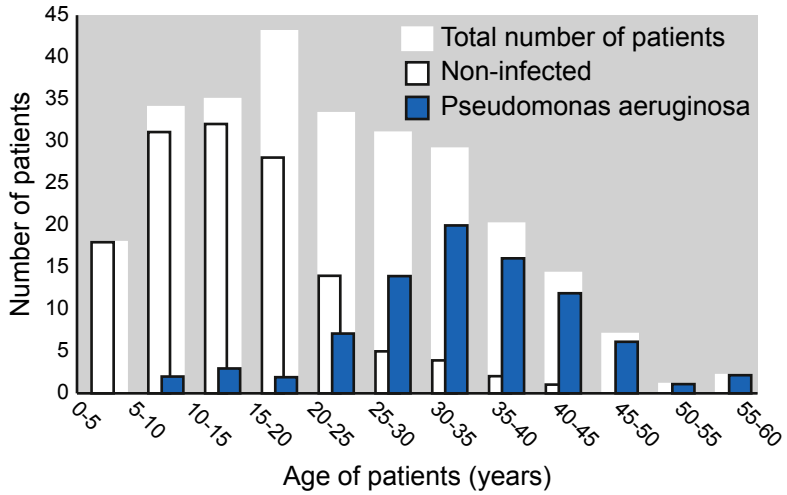


Figure 2: **Chronic Infection of CF Airways by *Pseudomonas aeruginosa* in 2008.** Number of CF paients in the Danish CF clinic who were registered as chronically infected with *P. aeruginosa* (blue) or not infected with a Gram-negative bacterium (white with black outline). Chronic infections by other Gram-negative species are not shown. Data provided by Tacjana Pressler, DMSc, Copenhagen CF database, 2008.

2.3.1 *CF and Laboratory Experiments*

When using an environment as the CF airways as a model system for evolution, it is valuable to make further investigations in controlled environments such as laboratory experiments. However, an optimal approach is to replicate the environment, where the evolution initially took place as much as possible, which can be difficult in the case of the CF model system[9].

In many studies of CF infections the phenotypic analyses are carried out in regular laboratory media[21, 22, 23, 24], with little resemblance to the mucus of the CF airway environmen t[9, 25, 26]. Not surprisingly, these differences in media composition and physical factors have been found to have a significant impact on phenotypic traits[9, 27, 28, 29].

However, we are slowly approaching a medium closely resembling the nutritional and physical aspects of CF sputum, called artificial sputum medium (ASM)[25, 26], showing similar phenotypic results for *P. aeruginosa* as found when using CF sputum as medium.

The introduction of ASM to the laboratory enables us to approach an experimental set-up more closely resembling the environment found in the CF sputum[26], and should further increase our knowledge

of the actual phenotypic adaptations found in the airways of CF patients.

2.3.2 *Environmental Heterogeneity*

The human airways are spatially heterogeneous, and can be roughly divided into the lower and upper airways, with the upper airways including the paranasal sinuses.

The sinuses are small air filled cavities situated in the cranium, with a different immune response and reduced antibiotic concentrations, compared to the lower airways. This is why they are considered to be a protected niche for bacteria, during airway infections of CF patients[30]. This will be addressed later in section 2.6 on page 19.

This spatial heterogeneity is predicted to result in adaptive radiation[4].

ADAPTIVE RADIATION is the adaptation of different sub-populations to different vacant niches. These could be the lungs or the sinuses of the human airways, where the different sub-populations become better equipped to proliferate in one environment (niche) compared to another.

Adaptive radiation has been shown in laboratory experiments involving *Pseudomonas fluorescens* in static cultures, where the stratification of the liquid media selected for three different phenotypic subpopulations[31]. It has also been shown for *P. aeruginosa* populations in CF, both by Markussen (2014)[21] and Jorth (2015)[32].

However, in an environment such as the CF airways, we have both chemical and physical temporal disturbances working against adaptive radiation, such as antibiotic treatments, physiotherapy, and perturbations. This results in a population that, *theoretically*, would evolve into GENERALISTS with a higher average fitness, instead of NICHE SPECIALISTS with a temporarily optimal fitness[4], Figure 3.

Despite what the theory dictates, the heterogeneity of bacterial populations in CF airway infections is evident from several studies of longitudinal single isolates, showing the presence of multiple contemporary sublineages. As an example, the phylogenetic reconstruction of 55 DK2 isolates from 21 Danish CF patients showed evidence of the following scenarios: Stable existence of a single clonal sublineage; stable coexistence of multiple independent sublineages, possibly due to niche differentiation; and sequential dominance of competing sub-

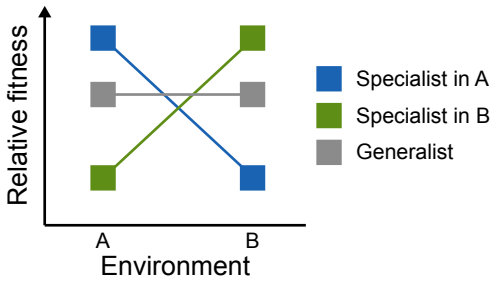


Figure 3: **Niche Specialisation.** Relative fitness for three populations, one which is adapted and specialised to environment A (blue), one which is adapted and specialised to environment B (green), and a generalist (grey), with a high average fitness.

lineages within the same patient[33].

NICHE DIFFERENTIATION is a process where natural selection drives organisms into different niches, either spatially or nutritionally. In other words, this process makes co-existence of different sublineages possible, by forcing them to use different resources and thereby hindering a diversity collapse caused by niche exclusion[34]. NICHE EXCLUSION is a process that results in a single genotype occupying the single niche provided by a perfectly homogeneous environment[35].

Niche differentiation is a process with which sympatric speciation can occur. SYMPATRIC SPECIATION is more commonly used to describe a specific speciation mechanism for sexually reproducing organisms, where different species evolve in overlapping or even identical ecological niches. In the example of DK2[33] it can be transferred to the generation of sub-speciation events seen in bacterial populations in CF.

The study of DK2 was based on longitudinally collected single isolates, compromising the depth with which population diversity could be examined. Examples where this has been investigated further, evidence of heterogeneous populations has been found for *P. aeruginosa*[36, 21, 37] and *Burkholderia dolosa*[38]. This will be discussed further in Section 2.4.3 on page 15.

Other mechanisms of diversity maintenance is the trade-offs that accompany the process of niche specialisation. This includes: MUTATION ACCUMULATION, where neutral mutations, that have accumulated in one environment, result in decreased fitness in another, and ANTAGONISTIC PLEITROPY, where a mutation creating a benefit in one environment is deleterious in another[3, 39]. In the case of mutation accumulation, an example could be the accumulation of mutations in genes involved in metabolic mechanisms. The possible

down regulation or destruction of some of these mechanisms might not cause decreased fitness in one environment, where alternative nutrients are present. But if the mutated organism is moved to an environment, where these mechanisms are necessary the accumulated mutations will be detrimental[39].

The difficulty of separating mutation accumulation from antagonistic pleitropy has been demonstrated for *E. coli* populations by Cooper (2000)[39]. Evidence of both mechanisms was found, but the decay of the catabolic mechanisms was often adaptive and could therefore not be solely attributed to random mutation accumulation. This caused them to attribute the functional decay to antagonistic pleitropy.

It is possible that mutation accumulation might be a significant factor when looking at populations with increased mutation rates, so called hypermutators, where deleterious and neutral mutations accumulate at a much higher rate than the adaptive beneficial mutations[40]. However, in the same *E. coli* study as mentioned above[39], no indication of this was found, when comparing populations with normal mutation rates and populations with increased mutation rates. It is argued that overall, the selective pressures and bottlenecks implicit in the study design had too high an impact for non-adaptive evolution to occur.

To my knowledge, there are no studies of CF bacterial populations where mutation accumulation has been found to be the primary cause of trade-off. This is probably because the CF airways pose a highly selective environment, where adaptive evolution is the front runner rather than random mutation accumulation, as shown above for the *E. coli* populations.

2.4 EVOLUTION AND ADAPTATION IN CF

In the moment a bacterium is moved from one environment to another, that being from soil to water, or from an external environment to the CF airways, adaptation is a necessity (Figure 4 A).
The adaptation to a new environment is guided by the selective pressures found in the environment (Figure 4 B) and natural selection, and it is through this process the transmission from being the cause of an intermittent colonisation to a chronic infection happens[41].

In general, adaptation can take two forms: phenotypic acclimation and genetic adaptation. PHENOTYPIC ACCLIMATION is based on alterations in gene regulatory mechanisms caused by environmental cues, such as toxins or nutrients, resulting in a phenotypic change

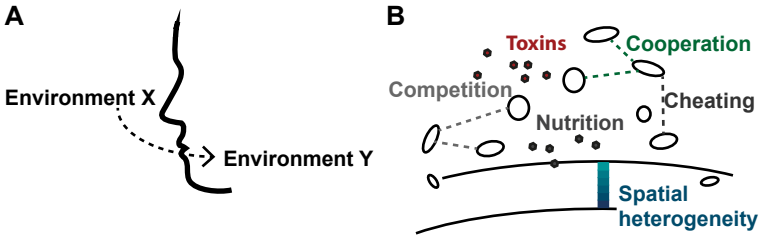


Figure 4: **Adaptation to a New Environment. A)** Transition from environment X to environment Y. **B)** Factors found in most environments to have an influence on adaptation and evolution.

without genotypic alterations. On the other hand, GENETIC ADAPTATION is facilitated by either acquisition of new genetic material from the environment, so called HORIZONTAL GENE TRANSFER (HGT), or changes to the pre-existing genetic content by mutation[42].

Phenotypic acclimation is probably one of the initial adaptive mechanisms, which together with genetic adaptation causes a rapid initial boost of phenotypic adaptation, as has been found in both laboratory experiments with *E. coli*[39] and in CF for *P. aeruginosa*[5]. This will be discussed in more detail in section: 2.6, subsection: 2.6.1 on page 22.

2.4.1 *Genetic Adaptation and Complex Evolutionary Trajectories*

In many cases, genetic adaptation has been found to cause increased antibiotic resistance[43, 44, 45, 46, 47, 48, 49].

Antibiotic resistance development is a well researched subject, especially in CF context because of the intense antibiotic treatment and the difficulties of eradication of *P. aeruginosa*[17]. Furthermore, antibiotic resistance is in general of great importance today because of the emerging threats of multi resistance, not only in *P. aeruginosa*, but in many other species as well. Therefore, we either need new drugs or we need to understand the evolution of resistance development, in order to improve treatment regimes and avoid resistance development when possible, and preferably both[50].

A specific example, where resistance development was linked to the introduction of a new drug was found by Marvig (2012)[43]. They were able to link the introduction of the antibiotic Azithromycin to the CF clinic in 2001, with the appearance of resistance generating mutations in the 23S rRNA of clinical isolates of *P. aeruginosa* from

CF patients.

Other cases of antibiotic resistance have also been found and linked to mutations in specific genes such as: *gyrA* and fluoroquinolone resistance, *nfxC* and quinolone resistance, and *nalD, mexA, mexB, mexR* and multi drug resistance[44, 45, 46, 47], etc.[48, 49].

In some cases multiple mutations are required for the development of resistance, some of which are deleterious when acting alone. Only when other mutations arise do they together result in a beneficial effect of resistance, subsiding the initial negative effects. One such example was found by Jochumsen (2013)[51]. They investigated the evolution of colistin resistance in *P. aeruginosa* and found an intricate web of molecular trajectories affected by antagonistic pleiotropy and EPISTATIC INTERACTIONS, i.e. non-additive effects caused by genetic interactions.

Another example of epistatic interactions has been shown by Damkiær (2013)[29] where mutations in four genes were found to be mutated in clinical isolates (*mucA, algU, rpoN*, and *lasR*). By reconstructing the mutated genes in the laboratory *P. aeruginosa* strain PAO1, they were able to show that mutations in either of the four genes alone or in combinations of up to three genes had no impact on resistance towards the two CF relevant antibiotics: ceftazidime and tobramycin. However, when all four genes were mutated in concert, they resulted in a significant increase of resistance to both antibiotics.

These complex trajectories can include points of temporarily lowered fitness compared to the ancestor, which was specifically shown in Jochumsen (2013)[51], where antibiotic resistance was used as a fitness indicator.

It is not uncommon in a complex environment that to go from one fitness peak to another it is necessary to cross a valley of lowered fitness (Figure 5). However the retainment of mutations with a deleterious effect is only possible in natural systems in two scenarios: One highly dependent on genetic drift and a large population size, and another relying on increased mutation rates[52, 40].

GENETIC DRIFT is the random change in frequency of genetic variants in a population over time. When natural selection is absent a genetic variant will, by chance events, go to fixation whereas other variants will be lost. The effect of genetic drift is dependent on population size, which can be exemplified by a coin toss: The chance of getting either heads or tails is equal, i.e. 0.5 for each. As the coin is tossed an infinite number of times the frequency with which you will

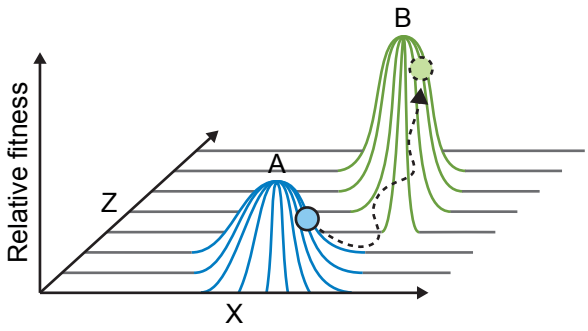


Figure 5: **Adaptive Peaks in a Fitness Landscape.** Demonstration of a population (circle) adapting to the different peaks of a fitness landscape. Peaks (A and B) represent optimal genotypes/phenotypes in a specific environment. To go from the current adaptive peak (A, and blue circle) the population has to cross a valley of lowered fitness (dotted line and arrow) to reach the higher fitness of the adaptive peak B (green circle). The axes X and Z are different genotypic or phenotypic dimensions that the population can move within to go towards different adaptive peaks, resulting in different relative fitness.

get heads will go towards 0.5. But, it is not unreasonable to think that if you only toss the coin 10 times you could end up with seven heads and three tails[52].

This entails that in the beginning of a colonisation, where the population is small, the chances of a non-beneficial mutation "sticking around" via genetic drift is low. On the other hand, in later stages, when the population is increasingly larger, rare mutations face a lower risk of being lost (the coin toss frequecy is close to 0.5) as a result of genetic drift and might therefore be able to "stick around" in the population long enough for another beneficial mutation to arise. An example of this could be the retention of mutations causing lowered fitness as compared to the ancestor, as found by Jochumsen (2013)[51].

Another way to get multiple rare mutations and traversing valleys of lowered fitness is by increasing the rate with which the population mutates. By increasing the mutation rate a higher number of mutations over a shorter period of time will automatically increase the chance of gaining rare mutations as well as the chance of gaining a subsequent beneficial mutation in time, – before the first mutation is lost from the population[40].

2.4.2 *Speeding up Evolution*

P. aeruginosa in CF, has been shown to have a mutation rate between 1.3 and 3.0 SNPs/year[21, 33, 53], for regular "non-mutator" populations. This rate can of course vary based on physiological factors such as mutagens or stress, and stress can be caused by the presence of antibiotics, which there is a lot of in the CF environment.

In hypermutator isolates sampled from two CF patients the mutation rate has been shown to be 50 and 106 SNPs/year[36], and in a simulation it has been shown that mutation rates up to 1000-fold higher than regular mutation rates in *E. coli* accelerate the adaptation rate significantly, without the mutation accumulation being detrimental for the population. The high mutation rate increases the number of beneficial mutations accumulating over time. However, neutral or deleterious mutations are accumulating in a much higher rate, thereby increasing the genetic load of the population at the same time[40].

Despite the increased genetic load, hypermutators have been found in a high number of CF patients with chronic *P. aeruginosa* infections[54, 55]. This resilience of hypermutator populations is probably caused by hitch-hiking, where the negative effects of the many deleterious or neutral mutations are balanced out by beneficial mutations[40].

More specifically, HITCH-HIKING is caused by the genetic link between genes or mutations causing them to be co-selected as a "linkage-group". In asexual populations, such as bacterial populations, the entire genome acts as a single linkage-group increasing the possibility of two mutations or genes to be co-selected, and thereby the possibility that one can hitch-hike with the other[3].

On several occasions hypermutation has been shown to cause increased antibiotic resistance in CF associated bacterial populations[54, 55, 56, 57], linking hypermutator generation to the antibiotic selection pressures found in CF patients. But as indicated by Fothergill (2007)[22], hypermutators do not always show increased antibiotic resistance, which is why antibiotic pressure cannot be the only driver of the generation of hypermutators.

Other factors that have been acknowledged as driving the persistence and occurrence of hypermutators are: oxidative stress[54], and the need for genotypic diversity, where increased mutation rates can increase the rate at which adaptive radiation can happen[36, 56].

The antibiotic pressure, the need to diversify, and a hypothesised small size of the bacterial populations in the initial colonisation period

is probably the reason why we identified hypermutators in young Danish CF patients (Study 2).

2.4.3 *Diversity and Resilience*

Diversity of a population will theoretically increase productivity and stability in an environment with temporal fluctuations, such as the CF airways. This principle is known as the INSURANCE HYPOTHESIS[58], and is supported by findings of a higher degree of resilience and productivity in highly diverse communities as compared to low diverse communities[59, 60]. This has been experimentally tested for biofilm populations of *P. aeruginosa* laboratory strains as well as CF derived strains, where increased diversity resulted in higher resilience towards physiological stress[61].

This mechanism is also known as BET-HEDGING, which is described as a stochastic switch between phenotypes with the aim of increasing overall fitness in multiple temporarily variable environments[62]. Under this name, the mechanism has been supported by experimental findings for *P. flourescens*, where a temporally alternating environment (static or shaken), resulted in mutations which in turn caused a phenotypic switch to increase the overall fitness of the population[63].

In this context, Lieberman (2014)[38] found that the population structure of *Burkholderia dolosa* in CF airways evolved according to a "diverse community model", with a high degree of diversity rather than a "dominant lineage model", where little diversity is seen. This finding, together with the evident resilience and persistence of *P. aeruginosa* in CF[36, 64, 21, 33, 23] despite heavy antibiotic treatment strategies[17], further supports the theory of persistence and resilience through diversity.

2.5 POPULATION HETEROGENEITY

The heterogeneity of the CF airways together with the theory of insurance effects and bet-hedging makes it highly plausible that the bacterial populations in the CF airways are heterogeneous and diverse[65]. This is also underlined by results from several studies, through which it has become increasingly obvious, that the populations correlated with CF infections are diverse[21, 36, 66, 67]. Correlating in a higher degree with the diverse community model than the dominant lineage model[38], as described above.

2.5.1 *Diverse Populations and Collections of Single Isolates*

In broad terms, diversity is upheld by a combination of ecological specialisation (i.e. antagonistic pleiotropy) and spatial segregation or spatial heterogeneity[4].

SPATIAL SEGREGATION is when physical barriers separate sub-populations and thereby hinders a collapse of diversity of the population as a whole. This is also a term found to describe speciation events of ALLOPATRIC SPECIATION, which to some degree can be transferred to the generation of sub-speciation events seen in bacterial populations of CF. In chronically infected CF patients spatial segregation has been found to maintain diversity by maintaining the infecting population in different compartments of the airways[21, 32]. It is also possible that this mechanism of allopatric speciation or spatial segregation was the initiator of the population segregation and sub-lineage formation.

The degree of spatial heterogeneity found in CF airways and the diversity found in the above mentioned studies, raise the question: Can we investigate the evolution and adaptation of diverse bacterial populations using collections of longitudinal *single* isolates?

Study 1:
Investigations of inter- and intra-clonal diversity of *Pseudomonas aeruginosa* populations in cystic fibrosis patients
Chapter 4 Section 4.1

To investigate this question we sequenced five metagenomes from sputum samples of four young CF patients, from whom we also had previous genome sequenced longitudinally collected single isolates.

We were able to link the clone type and the SNP profile of the single isolates to that of the metagenome(s) for each individual patient, even in the case where two patients were colonised by the same clone type as a result of patient-to-patient transmission.

We also analysed a patient with a hypermutator population by sampling two metagenomes with an interval of two weeks. From this we were able to show that even though the metagenomic approach increases the amount of information gained per sample it is still important, as it is for the single isolate approach, to have a comprehensive collection when

investigating evolution and population dynamics.

In the end we were able to conclude that the information gained by comprehensive collections of longitudinal single isolates is sufficient for the analysis of adaptation and evolution of *P. aeruginosa* to the CF airways.

As a further verification of the use of single isolates, it is important to note that some of the studies showing great diversity of the bacterial populations in CF were initially based on collections of single isolates[21, 36]. Indicating that it is possible to disclose diversity, as long as the collection of single isolates is comprehensive. By comprehensive, it is meant that, at the least, different compartments should be sampled and different sampling techniques should be used.

The two major studies showing spatial segregation of sub-populations are Markussen (2014)[21] and Jorth (2015)[32]. This segregation specifically hinders the use of single isolates for evolution studies, because all sub-populations are segregated and will not be identified by single isolate sampling. However, it should be noted that the diversity found in Markussen was initially identified by comprehensive sampling of single isolates from different sites of the airways (sinuses and lungs). It is also important to note the progressed state of the infection of the patients in both studies. Especially in Jorth (2015)[32], where the study is based on transplanted lungs. It is highly possible that the spatial segregation found to be important for diversity maintenance in these two studies primarily is a factor in progressed chronic infections. This segregation of sub-populations might not be relevant for the investigation of bacterial populations in CF patients at other colonisation/infection stages.

In younger, intermittently colonised patients or in patients with a less advanced chronic infection, mixing between airway compartments have been identified. Here the same genotypes and phenotypes were found in both the upper- and the lower-airways[30, 68, 69]. In these cases an alternative process of diversity maintenance could be clonal interference, which has been shown to occur and maintain diversity in populations of *E. coli* in a well mixed environment[70].

CLONAL INTERFERENCE is when multiple beneficial mutations occur in a population, at the same time, and thereby limit the spread of each other in the population. This competition between mutations will increase the time of fixation for any one of the mutations and thus slow down the overall rate of evolution and thereby maintain diversity[71].

Another way to maintain or create diversity is by negative-frequency-dependent selection. As with clonal interference, this has also been found in a homogeneous laboratory environment for *E. coli*[72]. Here, negative-frequency-dependent selection was shown to play a role in the maintenance of the mutationally derived diversity of the population.

2.5.2 *Frequency Dependent Selection and Cheater-Cooperator Dynamics*

NEGATIVE-FREQUENCY-DEPEDENT SELECTION is when a sub-lineage's fitness, in this case the ability to proliferate, is high when rare and low when abundant. An example, and what was found to be one of the drivers in the *E. coli* study[72], is cross-feeding: Here one sub-lineage (A) is dependent on substrates produced by another sub-lineage (B), so when A becomes abundant it out-competes B to a degree where B is incapable of producing enough substrate to sustain A. The frequency of A then decreases whereas B increases, until B is abundant enough to sustain A once more, Figure 6.

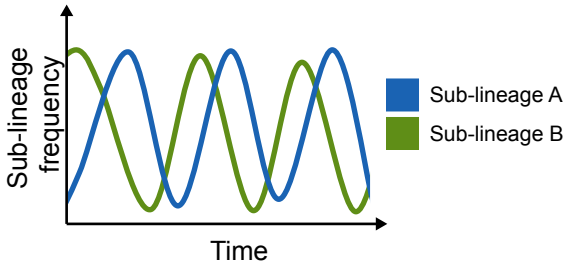


Figure 6: **Negative-Frequency-Dependent Selection.** A caricature of Negative-frequency-dependent selection, based on cross-feeding from sub-lineage B to sub-lineage A.

Cross feeding has also been indicated in *P. aeruginosa* populations from CF. Where Qin (2012)[73] found 19 of 27 collected isolates to be auxotrophic for at least one amino acid (argenine, lysine, methionine, alanine, and tryptophan), and all but one isolate could be complemented by at least one of the prototrophic isolates from the same patient.

Auxotrophy has also been shown for *Burkholderia cepacia*[74] and *S. aureus*[75] in CF, but as hypothesised by Barth (1995)[74] the auxotrophy might just be supported by the nutrient rich sputum found in CF, rather than cross-feeding.

Another type of frequency dependent selection can be found in the area of social microbiology of cheaters and cooperators.

Cooperators produce some kind of common good, for example iron scavenging agents, so called siderophores, that can facilitate the capture and uptake of iron in an environment with limited free iron, such as the CF airways. The cheaters are then isolates that do not have the cost of producing these agents, but lives through the siderophores produced by the cooperators. This diversity maintaining dynamic is however different from the negative-frequency-dependent selection, in that it is not maintained indefinitely. In most cases, the fitness of the cheaters will out-compete the cooperators, and the diversity will collapse.

In theory cheaters will have a fitness corresponding to a negative function of their frequency, and thereby be submitted to frequency dependent selection[76]. This dynamic has to my knowledge not been proven directly in CF airways. However, the frequency dependence of cheaters have been shown in mouse burn wound and chronic wound infection models[77], and the presence of cheaters and cooperators in young CF patients has recently been identified by Andersen (2015)[78]. Here siderophore cheats were identified to uptake iron by a receptor and only when cooperators were lost form the populations did the cheaters loose their receptors, possibly causing a diversity maintaining cheater-cooperator dynamic to break down.

One reason that frequency dynamics are difficult to observe is that it is necessary to have multiple isolates from each sample, to map the oscillations of the different sub-populations, and many studies of the general bacterial evolution in CF are based on single isolate collections[23, 33, 64, 78].

2.6 CLONAL PERSISTENCE AND ADAPTATION

As stated earlier, to follow the process of evolution and adaptation, it is valuable to be able to follow the same clonal lineage over time.

This is possible in CF infections because of a persistent clonal infection, enabling us to investigate the accumulation of mutations over the time of infection. It is important to note that even though we are talking about single clonal infections this does not preclude the evolution of intra-clonal diversity, and the diversity discussion from previously is equally applicable to single clone type and multi clone type infections. Furthermore, it will be evident from the following section that in CF we are in general talking about single clone type

infection scenarios.

Persistence of infection is specifically evident from the chronic infections shown for two Danish clonal lineages: DK1 and DK2, causing prolonged chronic infections for up to 32 and 35 years, respectively[21, 33]. This persistence of infection has also been shown by Smith (2006)[64] for a period of 8 years, by Cramer (2011)[23] for periods of 20 and 23 years in two German patients, and by Feliziani (2014)[36] for 6 and 20 years in an Argentinian and Danish patient, respectively.

The persistence of infection, despite the continuous antibiotic treatments, can be a result of three mechanisms: (1) transmission events between patients, (2) the presence of an internal protected reservoir within the patient, such as the sinuses, or (3) continous exposure to an environmental or hospital source[79, 80, 81, 82].

The first notion, of transmission, has been shown to happen on several occasions and was the instigating factor for the DK1 and DK2 spread among Danish CF patients[83, 84, 85, 86]. This patient-to-patient transmission will be discussed later in Section 2.8, on page 29.

The second notion, of an internal and protected reservoir, has been verified by studies of bacterial populations in the sinuses of CF patients. The sinuses are physically different from the lower airways with regards to sub-MIC antibiotic concentration (MIC: Minimum Inhibitory Concentration) and decreased immune system activity (IgA instead of IgG) as compared to the lungs. These factors make it highly likely that the sinuses can act as a protected niche, seeding the lower airways with adapted bacteria after or during antibiotic treatments[30, 68, 69, 87]. The main proof for the directionality of seeding from the sinuses to the lower airways is the fact that, despite employment of intense antibiotic treatment strategies, seemingly eradicating the lower airway bacterial infection[88, 17], infections keep re-occurring and persisting.

The third notion, of environmental sources, is thought to be the primary cause of the initial colonisation for Danich CF patients, and is also indicated by the findings of Study 2 in this thesis.

The general notion of the course of colonisation and infection in CF patients is as follows: An initial intermittent colonisation period with continued re-infections by environmental isolates. This period will extend through months or years, depending on the treatment and the adaptive state of the infecting bacteria[17, 89]. However, at some point the intermittent phase of colonisation will inevitably transition to a chronic infection, at which point there is little chance of eradication,

– if any[17] (Figure 7 A).

This persistence of the same clonal lineage in chronic CF patients is what makes it possible to follow the long term, within-patient, adaptation to the chronic CF environment. Contradictory to this is the hypothesised continuous re-colonisation by new environmental isolates in young CF patients. This would make it impossible to study the within-patient adaptation, because this adaptation would not exist.

Despite this notion of an initial phase of re-colonisations, the clonal persistence found in chronic CF patients has recently been identified in young Danish CF patients: A majority of 41 patients were shown to harbour the same clone type from the first culture of *P. aeruginosa* and until conclusion of the study [90].

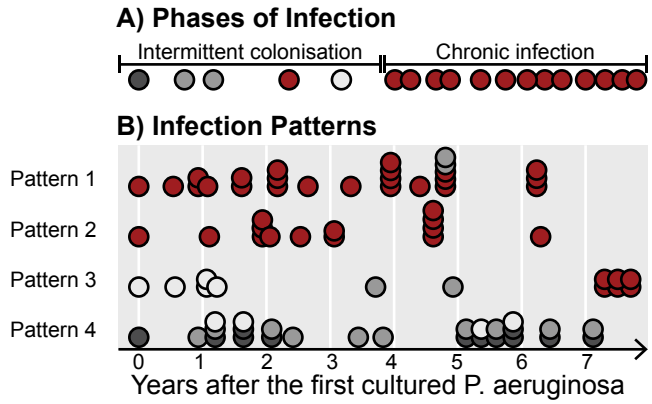


Figure 7: **Phases of infection and infection patterns.** Colours and shades of grey denote different clonal lineages. **(A)** The phases of infection, with an intermittent colonisation with environmental, unique clonal lineages and a subsequent chronic infection. **(B)** Four patterns of infection as is observed in the Danish CF children collection (Study 2). All four patterns are replicas of actual patterns found in four patients.

There are two current definitions of a chronic infection: (1) The Leeds criteria[91], where >50% of at least four samples collected over a year must be culture positive. (2) The Copenhagen criteria[92] where consecutive culture positive samples over a period of at least six months and/or an increased antibody response is required.

Currently, young Danish CF patients seem to harbour the same clonal lineage through many years, without being considered chronic under the above mentioned criteria. In this patient population we have identified four infection patterns (Figure 7 B), where the patterns of recurrent cultures of the same clonal lineage are seen in >75% of 41

patients[90], Figure 7 B patterns 1 and 2.

These results suggest that by the time the first culturing of *P. aeruginosa* is identified in the clinic, the patient has already transitioned from no colonisation to a chronic colonisation. However, in most cases this does not result in changes in the clinical conditions of the patients (personal communication with DMSc Helle Krogh Johansen).

These findings call for a re-evaluation of the current definitions of intermittent colonisation and chronic infection, as well as treatment regimes applied to a "first-time" infection of *P. aeruginosa*.

2.6.1 *Initial Infections and Pathoadaptation*

The rapid foothold, that the initial colonising lineage seems to take, is in clear accordance with the initial boost of phenotypic-adaptational changes that appears, when Bacteria move from one environment to another. This effect has been shown for *E. coli* in laboratory experiments[39], where a significant increase in fitness over the initial 20,000 generations was shown. It has also been shown for the phenotypic adaptation of *P. aeruginosa* collected from CF airways in the initial 50,000 generations[5] (Figure 8).

Study 2:
Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis
Chapter 4 Section 4.2

To investigate the initial phase of rapid adaptation we investigated the molecular evolution of 474 longitudinally collected single isolates from the airways of 34 young Danish CF patients, with an initial colonisation of *P. aeruginosa*.

We were able to identify 53 distinct clone types with inter clonal differences of >10,000 SNPs and intra clonal differences averaging on 122 SNPs (median = 9 SNPs). Of the 53 clone types ten were found to be harboured by multiple patients and of these, three were linked by epidemiological evidence of patient-to-patient transmission.

We identified 52 genes showing convergent molecular evolution among the clone types, and they corresponded to gene functions such as remodelling of regulatory networks and central metabolism, antibiotic resistance and virulence factors. *Further results are shown on page 26.*

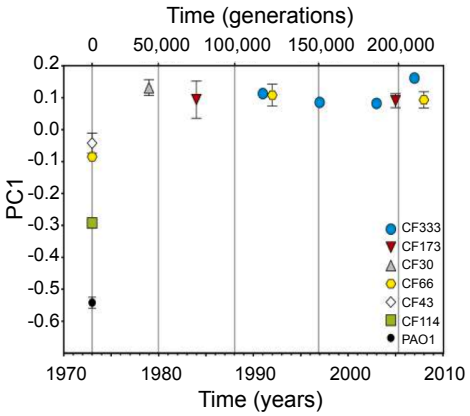


Figure 8: **The Initial Phenotypic Adaptational Boost.** Metabolic changes over 200,000 generations of *P. aeruginosa* evolution in the CF airways (PC1=63% of the total variation among isolates' metabolic capabilities). From Yang (2011)[5].

The 52 genes found in Study 2 are referred to in the paper as candidate "pathoadaptive".

PATHOADAPTATION was first used in 1999 by Sokurenko [41], to describe the adaptation of a micro-organism to a host environment through mutation of inherent genetic material. It does not include acquisition of virulence factors through HGT, but it includes both destruction and alteration of pre-existing functions.

In the context of CF, it has come to represent alterations significant to the adaptation to the airways, enabling the bacteria to become more resilient and persistent in the host environment.

We were able to link some of the clone types and pathoadaptive genes found in Study 2 to *P. aeruginosa* isolates from Italian CF patients (Study 3).

Study 3:
Within-host microevolution of *Pseudomonas aeruginosa* in Italian cystic fibrosis patients
Chapter 4 Section 4.3

For four Italian patients we tracked the adaptation and evolution over 19 years by sequencing and phenotypically analysing 26 *P. aeruginosa* isolates.

We were able to correlate genetic mutations with phenotypic changes and identify shared clone types between Italy and

Denmark, and show that clonal isolates within a country can be more distantly related than clonal isolates between countries.

In the investigation of pathoadaptive genes, we found an 11-fold enrichment of mutations in the 52 genes previously identified as pathoadaptive in Study 2 ($P(X \geqslant 34)\ pois(\lambda = 3.34) = 7.8x10-22$). Furthermore we found a signature of positive selection in these 52 genes, $dN/dS > 5$ (p=0.007) as compared to all other genes, $dN/dS = 0.7$. Overall, we identified a significant pattern of negative selection for two of the five clonal lineages (dN/dS = 0.7 and 0.3), whereas the others showed no significant pattern of positive or negative selection.

General conclusions of the extent to which natural selection has been the primary driving force of fixation of mutations, can be inferred by measuring relative rates of non-synonymous (dN) and synonymous (dS) mutations.

NON-SYNONYMOUS mutations have a direct impact on the mutated gene and can be contributed to missense and nonsense mutations, causing alterations of the amino acid sequence of encoded proteins, whereas SYNONYMOUS mutations do not cause these alterations of the amino acid sequence[71].

The dN/dS ratio is used to measure the degree with which natural selection is acting on the protein coding genome. POSITIVE SELECTION implies a ratio greater than one and indicates selection *for* mutation, whereas NEGATIVE SELECTION results in a ratio lower than one, and thus a selection for the *removal* of mutations, which is why it is also known as PURIFYING SELECTION.

Over the entire genome it is possible to have both positive and negative selection occurring at the same time[71]. This is the case in Study 3, where the general pattern of selection is either not significant for positive or negative selection or significant for negative selection. At the same time the selection pattern was found to be significant for positive selection in the 52 pathoadaptive genes previously identified in Study 2.

In general there are studies supporting both positive[64] and negative[79] selection as the driver of genomic evolution. In the end it is probably a combination of both, working at different positions on the genome, as was also indicated by Study 3.

2.6.2 *Convergent Genetic Evolution*

The study of pathoadaptation in CF has increased our knowledge of chronic infections by identifying common genetic and functional changes in the infecting bacteria, between clone types and between patients, also called CONVERGENT EVOLUTION. In CF the hallmarks of phenotypic adaptation of *P. aeruginosa* entails a decrease in growth rate, biofilm mode of growth, increased resistance towards antibiotics, and a loss of virulence factors[5, 21, 23, 24, 93, 94]. These phenotypic changes are all rooted in genotypic alterations and to some extent this phenotypic convergence is also reflected by a convergence of genotypic adaptations.

Looking at four large scale studies from Smith (2006)[64], Marvig (2013)[33], and Feliziani (2014)[36], together with Study 2 (Marvig *et al.*, Nat Genet), 19 candidate pathoadaptive genes were found in more than one study, Figure 9.

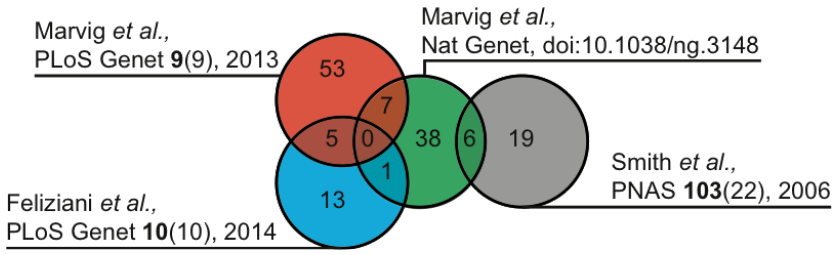


Figure 9: **Overlap of Pathoadaptive Genes from four Different Studies.** From Marvig (2015)[95].

The lack of overlap between the studies is to some extent a reflection of differences in patient populations, with regards to infection status (early/late stage colonisation/infection). It is also possibe that differences in treatment regimes, between countries and CF centres, play a role, as was hypothesised by Feliziani (2014)[36] to cause divergent genetic evolution of resistance, between a Danish and an Argentinian CF patient.

The impact of differences in the patient population, with regards to age and infection status is apparent, when looking closer at Study 2 and the study by Marvig (2013)[33], representing early colonisation and infection and progressed chronic infection by the DK2 clone, respectively. Three genes, found to be highly mutated in Study 2 (*mucA*, *algU*, and *mexZ*), were not included in the pathoadaptive gene list of DK2. Even though almost all isolates in the DK2 lineage have mutations in these genes. By looking back at the evolutionary history of DK2 we find that the genes were mutated early in the infection stage, and are therefore only represented by few unique mutations, thereby

excluding the genes from the pathoadaptive gene list, Figure 10.

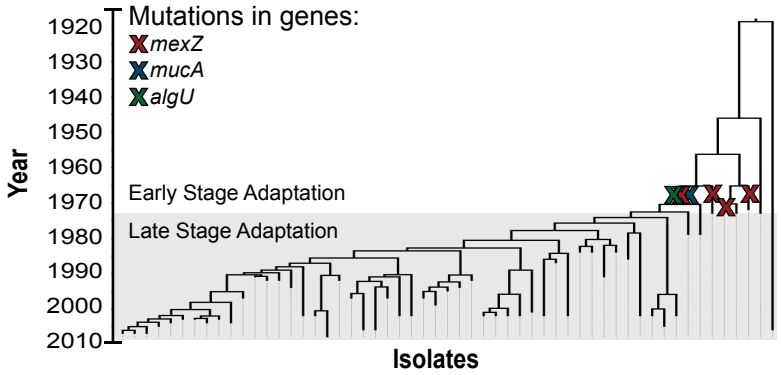


Figure 10: **Early and Late Stage Adaptation of DK2.** Bayesian phylogenetic reconstruction and divergence date estimates of the P. aeruginosa DK2 clones. X'es denote when mutations in each of three genes occured: red: *mexZ*, blue: *mucA*, and green: *algU*. Adapted from Marvig (2013)[33]

A third reason for the lack of overlap could be the approach with which the pathoadaptive genes are considered, where the genes are viewed as single entities instead of as a piece of a whole. This can be exemplified by the *retS-gacS-gacA-rsmZ-rsmA* system where, as explained further below, many genes can be hit and result in the same phenotypic adaptive change. However, in the approach used currently either one of the genes needs to be individually mutated enough times to be recognised as pathoadaptive.

One way of pursuing a more ideal approach, for the identification of pathoadaptive genes, could be to use a more holistic approach. Where as an example, the *retS-gacS-gacA-rsmZ-rsmA* system is recognised as a whole.

This could be achieved by methods such as those used in metabolic profiling, where genes are put into context of the pathways they are present in[96, 97].

2.7 HISTORIC CONTINGENCIES AND REGULATORY PATHWAYS

Study 2, continued:
Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis
Chapter 4 Section 4.2

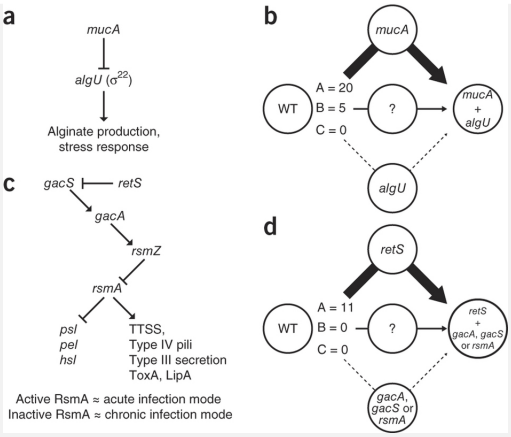


Figure 11: **Regulatory Pathways and Historic Contingencies.** The order of mutations in mutants with two nonsynonymous mutations in the same regulatory pathway. (a - d) We identified 25 and 11 unique double mutants with mutations in 2 of the genes in the mucA-algU (a) and retS-gacS-gacA-rsmZ-rsmA (c) regulatory pathways, respectively. On the basis of the phylogeny of the mutants, we were able to infer the order of the mutations in the mucA-algU (b) and retS-gacS-gacA-rsmZ-rsmA (d) regulatory pathways, that is, which of the possible mutational routes (A, B, or C) led to the double mutant. WT, wild type.

We were able to document historic contingencies of the two above mentioned systems of *mucA-algU* and *retS-gacS-gacA-rsmZ-rsmA*, Figure 11.

That is, we found a strong indication that the mutations in *algU* were dependent on a previous mutation in *mucA*, and the same pattern was found for *retS* and subsequent mutations in either of the genes *gacA/S* or *rsmA*.

These results are based on the investigation of pathways as a means to better understand the evolutionary processes. In particular the contingency of *retS-gacS-gacA-rsmZ-rsmA* illustrates the importance of this approach. Here we have an example where mutations in multiple genes are contingent on a previous mutation in another gene. However, either of the genes downstream of *retS* can result in the same phenotypic adaptation[98], making it difficult to identify the single genes as pathoadaptive, when looking at genetic markers of parallel evolution.

The above results also underline the importance of linking laboratory studies with findings in natural systems. The contingency of *mucA-algU* is in line with the phenotypic findings for *P. aeruginosa* isolates

collected from CF patients.

In general initial isolates, of CF airway infections, are non-mucoid. The transition to a chronic infection is often marked by a transition of the isolates to a mucoid phenotype, frequently caused by a *mucA* mutation[99, 100]. This is then abated by a subsequent mutation in for example *algU*[29, 94].

However, considering the model for *retS-gacS-gacA-rsmZ-rsmA* devised by Goodman (2004)[98] the results found in Study 2 indicate that by the first mutation in *retS* the population moves towards a chronic infection mode, but the subsequent mutations then re-direct the population towards an acute infection mode. This is in opposition to what we observe for the clone types found in the patients, which in reality move toward a chronic infection. This suggests that to consider the *retS-gacS-gacA-rsmZ-rsmA* signalling pathway as bimodal switch for acute or chronic infection, might be too simple.

By looking at the direct effects off the pathway, the chronic state equates transcription of the genes *psl, pel*, and *hsl*, where two of the genes (*psl* and *pel*) are responsible for the production of exopolysaccharides. Considering the decrease in attachment/biofilm formation, as seen in the laboratory, of CF isolates over time[24], this redirection away from transcription of *pel* and *psl* might actually not be contradictory.

This is not to say that *P. aeruginosa* does not form biofilm. On the contrary, it has been shown and is generally accepted, that they live in biofilm formations in the CF airways[101, 102].

However, considering the viscous mucus of CF patients, *P. aeruginosa* might not need to produce extracellular polysaccharides themselves, as they may be able to "cheat" off the human host, or cheat off other sub-populations.

It is highly unlikely that all isolates of the population mutate simultaneously, or that the mutations rapidly sweep the population, in line with the diversity found in most *P. aeruginosa* populations in CF patients. This population diversity could result in a cheater-cooperator dynamic reflecting a diversity of the population with regards to exopolysaccharide production, as have been shown previously for siderophore production.

We were able to find support for the initial hypothesis of the posibility that isolates cheat off the human host mucus, by investigating *P. aeruginosa* isolates from primary ciliary dyskinesia (PCD) patients.

PCD patients have normal (non-viscous) mucus compared to CF patients, but have similar infection patterns[103] also caused by mucus stasis in the airways. This is briefly explained below and in detail in Study 4.

Study 4:
Evolutionary and adaptational differences of *Pseudomonas aeruginosa* in primary ciliary dyskinesia and cystic fibrosis patients
Chapter 4 Section 4.4

By sequencing 35 longitudinal *P. aeruginosa* isolates and phenotypically characterising 41 from 12 chronically infected patients with primary ciliary dyskenesia (PCD) we identified 14 clone types. Five of which had previously been identified in CF patients in Study 2.

We were able to find overlaps of clone types and genetic adaptations previously identified in CF. Nonetheless, we were not able to find the same overlap of adaptive phenotypes. For the PCD isolates we identified biofilm formation to increase or remain status quo, over time. We did not find the rapid increase of generation time as found in CF isolates, and we found a retention of swimming motility and protease production. All opposing previous findings for CF isolates of chronic infections, and possibly a result of the low-viscocity mucus found in PCD as compared to CF.

2.8 COMPARATIVE CLONAL LINEAGES AND TRANSMISSION

In Study 2, 3, and 4 we identified clonal lineages found in multiple patients, between and within studies (Figure 12). This clonal overlap between different patients has been found before, with examples from the United Kingdom and the Liverpool Epidemic Strain (LES)[83], Australia and the Australian Epidemic Strain-2 (AES-2)[84], Denmark and the DK1 and DK2 clone types[85], amongst others[86]. And again between CF patients found in Denmark and Italy as well as Danish PCD patients (Study 2, 3, and 4, respectively).

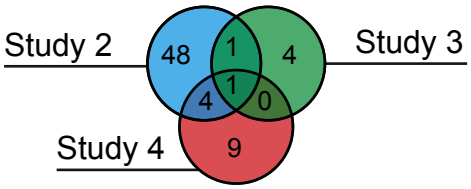


Figure 12: **Clone Type Overlap between Study 2, 3, and 4.** The numbers indicate unique clone types that have been identified in the individual or in multiple studies.

Similar events as described previously for the persistence of infection in CF, can also result in the presence of common clonal lineages: (1) The existence of common environmental clonal lineages, (2) indirect transmission via instruments in the clinic or exposure to the same environmental reservoirs, or (3) direct patient-to-patient transmission in the outpatient clinic or in the ward[79, 80].

The first notion is supported by the existence of common clonal lineages expanding to more than a single country as shown by the overlap of clone types between the patients from Denmark and Italy (Study 2 and 3, respectively) as well as the study by Wiehlmann (2007)[79], who identified a widespread geographic distribution of a few major clone types throughout Europe. The presence of DK06 in all three studies, without evidence of patient-to-patient transmission, further highlights that some clones are highly abundant in the environment. This is also shown in other studies, where the same clone type has been found in multiple patients[83, 85, 84, 86].

The second notion of common environmental reservoirs is supported by the overlap of clone types between patients within both Study 2 and Study 4, as well as the overlap between these studies. This is potentially caused by common environmental reservoirs since the PCD clinic and the CF clinic resides at the same hospital.

The third and last notion, of patient-to-patient transmission is supported by studies of transmission of *Burkholderia cepacia*[104] and *P. aeruginosa*[5, 33, 86] including Study 2. As well as findings of viable Gram negative bacteria that can travel up to 4 m and persist for 45 min in cough aerosols[81, 82]. This risk of transmission of already adapted clonal lineages to uninfected patients has lead to patient segregation in the Copenhagen CF clinic, since the early eighties[105]. Greatly minimising the risk of a chronic patient transferring an already adapted clonal lineage to an uninfected patient, or to a patient infected with a naive environmental clonal lineage[15, 89].

3

CONCLUSIONS AND FUTURE PERSPECTIVES

3.1 CONCLUSIONS

Genomic investigations of longitudinally collected isolates and metagenomes have provided evidence for the heterogeneity of infecting bacterial populations in CF airways. This diversity of the infecting populations questions the use of single longitudinal isolates as a means to elucidate the evolution and adaptation of an entire bactarial population.

We sought to investigate this by analysing five sputum metagenomes together with longitudinally collected single isolates from four CF patients in Study 1. With this approach we were able to link the clone type and the SNP profiles of the single isolates to that of the metagenome(s) for each individual patient. This indicates that comprehensive sampling of single isolates is sufficient for the study of evolution and adaptation of *P. aeruginosa* in CF airways. Another important finding was that even though the metagenomic approach increases the amount of information gained per sample, it is still important to have a comprehensive collection when investigating evolution and population dynamics, – as is the case for the single isolate approach.

By the genome sequencing of longitudinal single isolates from young Danish CF patients, we found that 75% of the patients harboured the same clone type from the first culturing of *P. aeruginosa* an onward. This calls for a re-evaluation of the current definitions of intermittent colonisation and chronic infection, as well as the treatment regimes applied to a "first-time" infection.

Nevertheless this persistent infection with the same clonal lineage enabled us to investigate the initial phases of adaptation found in CF airways.

By genome sequencing a total of 535 isolates collected from 34 young Danish CF patients (474 isolates, Study 2), four Italian CF patients (26 isolates, Study 3) and 12 PCD patients (35 isolates, Study 4) we were able to identify initial genetic adaptation markers (pathoadaptive genes).

The genome comparisons of the 474 single isolates, constituting 53 unique clone types, revealed a high degree of convergence, where we

identified 52 candidate pathoadaptive genes. These were later identified to be significant for adaptation of *P. aeruginosa* in Italian CF paients as well, and overlapping with mutated genes found to be significant for the adaptation to PCD airways.

This list of genes suggests a role in host adaptation for remodelling of regulatory networks and central metabolism, acquisition of antibiotic resistance and loss of extracellular virulence factors. This is consistent with both genomic and phenotypic findings of other studies investigating *P. aeruginosa* infections in CF patients[15, 21, 36, 64, 96, 97, 106].

Furthermore, by investigating mutation patterns in key regulatory networks, we were able to identify a dependence of sequential mutations. Downstream transcriptional regulators were found to be contingent on mutations in upstream regulators, further supporting the importance of remodelling regulatory networks for adaptation to CF airways.

By comparison to three other large scale studies, Smith (2006)[64], Marvig (2013)[33], and Feliziani (2014)[36] we found merely 19 genes overlapping at least two studies, and none overlapping all four studies. This poses an important question: If these candidate pathoadaptive genes are important for the adaptation to the CF airways in general, then why do they not overlap?

One reason to look for these genes in the first place, was to find specific markers for the transition from colonisation to infection, but if there is no overlap between studies then this is not possible.

There is ample evidence that the candidate pathoadaptive genes are a result of convergent evolution within the studies, however not so much between the studies.

By looking closer at the patient groups represented in the studies it was shown that the infection stage of the isolates investigated was important to which genes were identified as highly mutated. As it turns out, some genes are primarily mutated in the early stage adaptation, and therefore possibly not included in the list of candidate pathoadaptive genes in studies of late stage adaptation.

By drawing a parallel to the historic contingencies found in Study 2 of the regulatory pathway of *retS-gacS-gacA-rsmZ-rsmA*, it is possible that by changing the approach from looking at single genes to looking at pathways, and thereby groups of genes we might find a higher degree of overlap.

In Study 2 we also found evidence of patient-to-patient transmission, as have been found in many studies previously[83, 85, 84, 86]. We were able to link possible transmission events to temporary overlaps of hospital visits, providing epidemiological evidence to the genomic results in three cases.

By comparing the sequencing results from the young CF patients' *P. aeruginosa* isolates, to results found for Italian CF patients we were able to correlate and identify convergence of genetic adaptation across two countries. We also found that clone types shared between patients can show a greater similarity between countries than within, underlining the presence of common clonal lineages across borders. Clone types found in either or both studies were also found in PCD patients, indicating that there are no specific clone types responsible for infection of either of these patient groups.

By an investigation of both genetic and phenotypic adaptational markers for the PCD isolates we identified a discontinuity in comparison to CF isolates: In other words, we identified candidate pathoadaptive genes for the PCD isolates, which correspond with previous findings of CF isolates. However, despite this overlap of genetic adaptation we did not identify a similar overlap in phenotypic adaptation. We identified retention of swimming motility, protease production and attachment, as a contrast to what has been found in CF.

These results indicate that the viscosity of the mucus provides a significant selective pressure on the infecting *P. aeruginosa* populations, which is in agreement with the structural differences of CF mucus as compared with normal mucus[14].

Independently and together the studies presented in this thesis provide new knowledge of adaptation and evolution in both CF and PCD airways. With further characterisation of pathoadaptive mutations and phenotypic adaptations it should be possible to translate these results into clinically relevant information, leading to better epidemiological predictions, valuable information with regards to treatment strategies, and perhaps extrapolation of this knowledge to other infection scenarios.

OVER ALL: Through the convergence of genetic and phenotypic adaptations observed in CF studies and by linking processes of evolution to these observations, it is shown that collections of *P. aeruginosa* isolates from CF patients provide a valuable basis for the study of adaptation and evolution in natural environments.

3.2 FUTURE PERSPECTIVES

In Study 2 we found that adaptation was associated with mutations in regulatory networks, which probably changes the functionality of these networks and guides adaptation towards specific evolutionary routes. However, despite the high number of sequenced isolates and clone types presented here, we have still not been able to pinpoint any specific evolutionary route, even though the phenotypic results point toward a specific end point. A possible explanation for the lack of a more clear genetic adaptational route might be patient specific factors that are as yet not fully incorporated in the analyses of adaptation and evolution. These factors include co-infection or co-colonisation of multiple species or multiple clones, or differentiated antibiotic treatment.

Since co-colonisation has already been established to have an effect on the evolution of both *P. aeruginosa* and *S. aureus*[28, 107]. A logical next step could be to integrate the knowledge of co-infections and inter-species dynamics to get a better understanding of the microbial evolution. By concurrently addressing the limitations of resolution when using single isolates, this could be carried out by metagnomic sampling in combination with single isolates of the species of particular interest.

The antibiotic treatments could be incorporated to a larger extent than is the case at the moment. In general antibiotic treatments are only incorporated by taking the drugs used against the specific bacteria of interest into account. However, because of the possibility of interspecies interactions, drugs that do not have a *direct* effect on the bacterial organism of interest may still have an unrecognised potential as an *indirect* selective factor for adaptation and evolution.

Another approach could be the application of methods previously used for metabolic modelling, where genes are not considered as separate entities but rather as a part of different pathways. This, may give us an opportunity to connect genes in an until now unknown pattern of interrelated highways of adaptation and evolution, giving us the possibility to more thoroughly understand evolution, as well as increase our chances of identifying key markers of transition from an initial environmental isolate to persistent colonisation and further to chronic infection. In turn, this may enable the development of better diagnostic and treatment approaches and postpone or even eliminate the final progression to a chronic and incurable bacterial infection with *P. aeruginosa*.

BIBLIOGRAPHY

[1] M. J. Mandel, M. S. Wollenberg, E. V. Stabb, K. L. Visick, and E. G. Ruby. A single regulatory gene is sufficient to alter bacterial host range. *Nature*, 458(7235):215–8, Mar 2009.

[2] I. Lozada-Chávez, S. C. Janga, and J. Collado-Vides. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Research*, 34(12):3434–45, 2006.

[3] S. F. Elena and R. E. Lenski. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, 4(6):457–69, Jun 2003.

[4] R. Kassen and P. B. Rainey. The ecology and genetics of microbial diversity. *Annual Review of Microbiology*, 58:207–231, May 2004.

[5] L. Yang, L. Jelsbak, R. L. Marvig, S. Damkiær, C. T. Workman, M. H. Rau, S. K. Hansen, A. Folkesson, H. K. Johansen, O. Ciofu, N. Høiby, M. O. A. Sommer, and S. Molin. Evolutionary dynamics of bacteria in a human host environment. *Proceedings of the National Academy of Science USA*, 108(18):7481–6, May 2011.

[6] C. Fraser, W. P. Hanage, and B. G. Spratt. Recombination and the nature of bacterial speciation. *Science*, 315(5811):476–80, Jan 2007.

[7] A. Jeltsch. Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene*, 317(1-2):13–6, Oct 2003.

[8] M. C. Horner-Devine, K. M. Carney, and B. J. M. Bohannan. An ecological perspective on bacterial biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, 271(1535):113–22, Jan 2004.

[9] A. E. L. Roberts, K. N. Kragh, T. Bjarnsholt, and S. P. Diggle. The limitations of in vitro experimentation in understanding biofilms and chronic infection. *Journal of Molecular Biology*, a head of print, Sep 2015.

[10] R. Maddamsetti, R. E. Lenski, and J. E. Barrick. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with escherichia coli. *Genetics*, 200(2):619–31, Jun 2015.

[11] J. R. Riordan, J. M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic, and J. L. Chou. Identification of the cystic fibrosis gene: cloning and characterization of complementary dna. *Science*, 245(4922):1066–73, Sep 1989.

[12] M. J. Welsh and A. E. Smith. Molecular mechanisms of cftr chloride channel dysfunction in cystic fibrosis. *Cell*, 73(7):1251–4, Jul 1993.

[13] R. C. Boucher. Relationship of airway epithelial ion transport to chronic bronchitis. *Proceedings of the American Thoracic Society*, 1(1):66–70, 2004.

[14] D. Borowitz. Cftr, bicarbonate, and the pathophysiology of cystic fibrosis. *Pediatric Pulmonology*, 50 Suppl 40:2S4–S30, Oct 2015.

[15] A. Folkesson, L. Jelsbak, L. Yang, H. K. Johansen, O. Ciofu, N. Høiby, and S. Molin. Adaptation of pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. *Nature Reviews Microbiology*, 10(12):841–51, Dec 2012.

[16] E. L. Salsgiver, A. K. Fink, E. A. Knapp, J. J. LiPuma, K. N. Olivier, B. C. Marshall, and L. Saiman. Changing epidemiology of the respiratory bacteriology of patients with cystic fibrosis. *Chest*, Jul 2015.

[17] N. Høiby, B. Frederiksen, and T. Pressler. Eradication of early pseudomonas aeruginosa infection. *Journal of Cystic Fibrosis*, 4 Suppl 2:49–54, Aug 2005.

[18] T. Bjarnsholt, K. Kirketerp-Møller, P. Ø. Jensen, K. G. Madsen, R. Phipps, K. Krogfelt, N. Høiby, and M. Givskov. Why chronic wounds will not heal: a novel hypothesis. *Wound Repair and Regeneration*, 16(1):2–10, 2008.

[19] M. D. Obritsch, D. N. Fish, R. MacLaren, and R. Jung. Nosocomial infections due to multidrug-resistant pseudomonas aeruginosa: epidemiology and treatment options. *Pharmacotherapy*, 25(10):1353–64, Oct 2005.

[20] D. M. Livermore. Multiple mechanisms of antimicrobial resistance in pseudomonas aeruginosa: our worst nightmare? *Clinical Infectious Dieseases*, 34(5):634–40, Mar 2002.

[21] T. Markussen, R. L. Marvig, M. Gómez-Lozano, K. Aanæs, A. E. Burleigh, N. Høiby, H. K. Johansen, S. Molin, and L. Jelsbak. Environmental heterogeneity drives within-host diversification and evolution of pseudomonas aeruginosa. *MBio*, 5(5):e01592–14, 2014.

[22] J. L. Fothergill, S. Panagea, C. A. Hart, M. J. Walshaw, T. L. Pitt, and C. Winstanley. Widespread pyocyanin over-production among isolates of a cystic fibrosis epidemic strain. *BMC Microbiology*, 7:45, 2007.

[23] N. Cramer, J. Klockgether, K. Wrasman, M. Schmidt, C. F. Davenport, and B. Tümmler. Microevolution of the major common pseudomonas aeruginosa clones c and pa14 in cystic fibrosis lungs. *Environmental Microbiology*, 13(7):1690–704, Jul 2011.

[24] B. Lee, J. A. J. Haagensen, O. Ciofu, J. B. Andersen, N. Høiby, and S. Molin. Heterogeneity of biofilms formed by nonmucoid pseudomonas aeruginosa isolates from patients with cystic fibrosis. *Journal of Clinical Microbiology*, 43(10):5247–55, Oct 2005.

[25] S. Kirchner, J. L. Fothergill, E. A. Wright, C. E. James, E. Mowat, and C. Winstanley. Use of artificial sputum medium to test antibiotic efficacy against pseudomonas aeruginosa in conditions more relevant to the cystic fibrosis lung. *Journal of Visualized Experiments*, (64):e3857, 2012.

[26] K. L. Palmer, L. M. Aye, and M. Whiteley. Nutritional cues control pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum. *Journal of Bacteriology*, 189(22):8079–87, Nov 2007.

[27] J. A. J. Haagensen, D. Verotta, L. Huang, A. Spormann, and K. Yang. New in vitro model to study the effect of human simulated antibiotic concentrations on bacterial biofilms. *Antimicrobial Agents and Chemotherapy*, 59(7):4074–81, Jul 2015.

[28] C. F. Michelsen, A-M. J. Christensen, M. S. Bojer, N. Høiby, H. Ingmer, and L. Jelsbak. Staphylococcus aureus alters growth activity, autolysis, and antibiotic tolerance in a human host-adapted pseudomonas aeruginosa lineage. *Journal of Bacteriology*, 196(22):3903–11, Nov 2014.

[29] S. Damkiær, L. Yang, S. Molin, and L. Jelsbak. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proceedings of the National Academy of Science USA*, 110(19):7766–71, May 2013.

[30] H. K. Johansen, K. Aanaes, T. Pressler, K. G. Nielsen, J. Fisker, M. Skov, N. Høiby, and C. von Buchwald. Colonisation and infection of the paranasal sinuses in cystic fibrosis patients is accompanied by a reduced pmn response. *Journal of Cystic Fibrosis*, 11(6):525–31, Dec 2012.

[31] P. B. Rainey and M. Travisano. Adaptive radiation in a heterogeneous environment. *Nature*, 394(6688):69–72, Jul 1998.

[32] P. Jorth, B. J. Staudinger, X. Wu, K. Hisert, H. Hayden, J. Garudathri, C. L. Harding, M. C. Radey, A. Rezayat, G. Bautista, W. R. Berrington, A. F. Goddard, C. Zheng, A. Angermeyer, M. J. Brittnacher, J. Kitzman, J. Shendure, C. L. Fligner, J. Mittler, M. L. Aitken, C. Manoil, J. E. Bruce, T. L. Yahr, and P. K. Singh. Regional isolation drives bacterial diversification within cystic fibrosis lungs. *Cell Host & Microbe*, 18:1–12, Sep 2015.

[33] R. L. Marvig, H. K. Johansen, S. Molin, and L. Jelsbak. Genome analysis of a transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genetics*, 9(9):e1003741, 2013.

[34] R. MacArthur and R. Levins. The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, 101(921):377–385, 1967.

[35] G. Hardin. The competitive exclusion principle. *Science*, 131(3409):1292–1297, Apr 1960.

[36] S. Feliziani, R. L. Marvig, A. M. Luján, A. J. Moyano, J. A. Di Rienzo, H. K. Johansen, S. Molin, and A. M. Smania. Coexistence and within-host evolution of diversified lineages of hypermutable pseudomonas aeruginosa in long-term cystic fibrosis infections. *PLoS Genetics*, 10(10):e1004651, Oct 2014.

[37] J. Diaz Caballero, S. T. Clark, B. Coburn, Y. Zhang, P. W. Wang, S. L. Donaldson, D. E. Tullis, Y. C. W. Yau, V. J. Waters, D. M. Hwang, and D. S. Guttman. Selective sweeps and parallel pathoadaptation drive pseudomonas aeruginosa evolution in the cystic fibrosis lung. *MBio*, 6(5), 2015.

[38] T. D. Lieberman, K. B. Flett, I. Yelin, T. R. Martin, A. J. McAdam, G. P. Priebe, and R. Kishony. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nature Genetics*, 46(1):82–7, Jan 2014.

[39] V. S. Cooper and R. E. Lenski. The population genetics of ecological specialization in evolving escherichia coli populations. *Nature*, 407(6805):736–9, Oct 2000.

[40] F. Taddei, M. Radman, J. Maynard-Smith, B. Toupance, P. H. Gouyon, and B. Godelle. Role of mutator alleles in adaptive evolution. *Nature*, 387(6634):700–2, Jun 1997.

[41] E. V. Sokurenko, D. L. Hasty, and D. E. Dykhuizen. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends in Microbiology*, 7(5):191–5, May 1999.

[42] P. B. Rainey. Bacterial populations adapt - genetically, by natural selection - even in the lab! *Microbiology Today*, 31:160 – 162, 2004.

[43] R. L. Marvig, M. S. R. Søndergaard, S. Damkiær, N. Høiby, H. K. Johansen, S. Molin, and L. Jelsbak. Mutations in 23s rrna confer resistance against azithromycin in pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, 56(8):4519–21, Aug 2012.

[44] M. L. Sobel, D. Hocquet, L. Cao, P. Plesiat, and K. Poole. Mutations in pa3574 (nald) lead to increased mexab-oprm expression and multidrug resistance in laboratory and clinical isolates of pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, 49(5):1782–6, May 2005.

[45] P. G. Higgins, A. C. Fluit, D. Milatovic, J. Verhoef, and F. J. Schmitz. Mutations in gyra, parc, mexr and nfxb in clinical isolates of pseudomonas aeruginosa. *International Journal of Antimicrobial Agents*, 21(5):409–13, May 2003.

[46] I. Ziha-Zarifi, C. Llanes, T. Köhler, J. C. Pechere, and P. Plesiat. In vivo emergence of multidrug-resistant mutants of pseudomonas aeruginosa overexpressing the active efflux system mexa-mexb-oprm. *Antimicrobial Agents and Chemotherapy*, 43(2):287–91, Feb 1999.

[47] H. Fukuda, M. Hosaka, S. Iyobe, N. Gotoh, T. Nishino, and K. Hirai. nfxc-type quinolone resistance in a clinical isolate of pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, 39(3):790–792, Mar 1995.

[48] T. Strateva and D. Yordanov. Pseudomonas aeruginosa - a phenomenon of bacterial resistance. *Jounal of Medical Microbiology*, 58(Pt 9):1133–48, Sep 2009.

[49] K. Poole. Multidrug efflux pumps and antimicrobial resistance in pseudomonas aeruginosa and related organisms. *Journal of Molecular Microbiology and Biotechnology*, 3(2):255–264, 2001.

[50] G. Dantas and M. O. A. Sommer. How to fight back against antibiotic resistance. *American Scientist*, 102:42 – 51, 2014.

[51] N. Jochumsen. *The Causes and Consequences of Antibiotic Resistance Evolution in Microbial Pathogens*. PhD thesis, Technical University of Denmark, Anker Engelunds vej 1, 2800 Lyngby, 2013.

[52] J. Masel. Genetic drift. *Current Biology*, 21(20):R837–8, Oct 2011.

[53] L. A. Snyder, N. J. Loman, L. A. Faraj, K. Levi, G. Weinstock, T. C. Boswell, M. J. Pallen, and D. A. Ala'Aldeen. Epidemiological investigation of pseudomonas aeruginosa isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing. *Eurosurveillance*, 18(42), 2013.

[54] O. Ciofu, B. Riis, T. Pressler, H. E. Poulsen, and N. Høiby. Occurrence of hypermutable pseudomonas aeruginosa in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. *Antimicrobial Agents and Chemotherapy*, 49(6):2276–82, Jun 2005.

[55] A. Oliver, R. Cantón, P. Campo, F. Baquero, and J. Blázquez. High frequency of hypermutable pseudomonas aeruginosa in cystic fibrosis lung infection. *Science*, 288(5469):1251–4, May 2000.

[56] O. Ciofu, L. F. Mandsberg, T. Bjarnsholt, T. Wassermann, and N. Høiby. Genetic adaptation of pseudomonas aeruginosa during chronic lung infection of patients with cystic fibrosis: strong and weak mutators with heterogeneous genetic backgrounds emerge in muca and/or lasr mutants. *Microbiology*, 156(Pt 4):1108–19, Apr 2010.

[57] A. Ferroni, D. Guillemot, K. Moumile, C. Bernede, M. Le Bourgeois, S. Waernessyckle, P. Descamps, I. Sermet-Gaudelus, G. Lenoir, P. Berche, and F. Taddei. Effect of mutator p. aeruginosa on antibiotic resistance acquisition and respiratory function in cystic fibrosis. *Pediatric Pulmonology*, 44(8):820–5, Aug 2009.

[58] S. Yachi and M. Loreau. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proceedings of the National Academy of Science USA*, 96:1463–1468, Feb 1999.

[59] D. Tilman, D. Wedin, and J. Knops. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature*, 379(6567):718–720, Feb 1996.

[60] D. Tilman, J. Hill, and C. Lehman. Carbon-negative biofuels from low-input high-diversity grassland biomass. *Science*, 314(5805):1598–600, Dec 2006.

[61] B. R. Boles, M. Thoendel, and P. K. Singh. Self-generated diversity produces "insurance effects" in biofilm communities. *Proceedings of the National Academy of Science USA*, 101(47):16630–5, Nov 2004.

[62] J-W. Veening, W. K. Smits, and O. P. Kuipers. Bistability, epigenetics, and bet-hedging in bacteria. *Annual Review of Microbiology*, 62:193–210, 2008.

[63] H. J. E. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey. Experimental evolution of bet hedging. *Nature*, 462(7269):90–3, Nov 2009.

[64] E. E. Smith, D. G. Buckley, Z. Wu, C. Saenphimmachak, L. R. Hoffman, D. A. D'Argenio, S. I. Miller, B. W. Ramsey, D. P. Speert, S. M. Moskowitz, J. L. Burns, R. Kaul, and M. V. Olson. Genetic adaptation by pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proceedings of the National Academy of Science USA*, 103(22):8487–92, May 2006.

[65] D. Nguyen and P. K. Singh. Evolving stealth: genetic adaptation of pseudomonas aeruginosa during cystic fibrosis infections. *Proceedings of the National Academy of Science USA*, 103(22):8305–6, May 2006.

[66] E. Mowat, S. Paterson, J. L. Fothergill, E. A. Wright, M. J. Ledson, M. J. Walshaw, M. A. Brockhurst, and C. Winstanley. Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *American Journal of Respiratory and Critical Care Medicine*, 183(12):1674–9, Jun 2011.

[67] M. L. Workentine, C. D. Sibley, B. Glezerson, S. Purighalla, J. C. Norgaard-Gron, M. D. Parkins, H. R. Rabin, and M. G. Surette. Phenotypic heterogeneity of pseudomonas aeruginosa populations in a cystic fibrosis patient. *PLoS One*, 8(4):e60225, 2013.

[68] S. K. Hansen, M. H. Rau, H. K. Johansen, O. Ciofu, L. Jelsbak, L. Yang, A. Folkesson, H. Ø. Jarmer, K. Aanæs, C. von Buchwald, N. Høiby, and S. Molin. Evolution and diversification of pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have implications for chronic lung infection. *The ISME Journal*, 6(1):31–45, Jan 2012.

[69] J. G. Mainz, L. Naehrlich, M. Schien, M. Käding, I. Schiller, S. Mayr, G. Schneider, B. Wiedemann, L. Wiehlmann, N. Cramer, W. Pfister, B. C. Kahl, J. F. Beck, and B. Tümmler. Concordant genotype of upper and lower airways p aeruginosa and s aureus isolates in cystic fibrosis. *Thorax*, 64(6):535–40, Jun 2009.

[70] J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. Genome evolution and adaptation in a long-term experiment with escherichia coli. *Nature*, 461(7268):1243–7, Oct 2009.

[71] M. K. Burke. How does adaptation sweep through the genome? insights from long-term selection experiments. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749):5029–38, Dec 2012.

[72] D. E. Rozen and R. E. Lenski. Long-term experimental evolution in escherichia coli. viii. dynamics of a balanced polymorphism. *The American Society of Naturalists*, 155(1):24–35, Jan 2000.

[73] X. Qin, D. M. Zerr, M. A. McNutt, J. E. Berry, J. L. Burns, and R. P. Kapur. Pseudomonas aeruginosa syntrophy in chronically colonized airways of cystic fibrosis patients. *Antimicrobial Agents and Chemotherapy*, 56(11):5971–81, Nov 2012.

[74] A. L. Barth and T. L. Pitt. Auxotrophy of burkholderia (pseudomonas) cepacia from cystic fibrosis patients. *Journal of Clinical Microbiology*, 33(8):2192–4, Aug 1995.

[75] B. Kahl, M. Herrmann, A. S. Everding, H. G. Koch, K. Becker, E. Harms, R. A. Proctor, and G. Peters. Persistent infection with small colony variant strains of staphylococcus aureus in patients with cystic fibrosis. *Journal of Infectious Diseases*, 177(4):1023–9, Apr 1998.

[76] F. Harrison, L. E. Browning, M. Vos, and A. Buckling. Cooperation and virulence in acute pseudomonas aeruginosa infections. *BMC Biology*, 4:21, 2006.

[77] K. P. Rumbaugh, S. P. Diggle, C. M. Watters, A. Ross-Gillespie, A. S. Griffin, and S. A. West. Quorum sensing and the social evolution of bacterial virulence. *Current Biology*, 19(4):341–5, Feb 2009.

[78] S. B. Andersen, R. L. Marvig, S. Molin, H. K. Johansen, and A. S. Griffin. Long-term social dynamics drive loss of function in pathogenic bacteria. *Proceedings of the National Academy of Science USA*, 112(34):10756–61, Aug 2015.

[79] L. Wiehlmann, G. Wagner, N. Cramer, B. Siebert, P. Gudowius, G. Morales, T. Köhler, C. van Delden, C. Weinel, P. Slickers, and B. Tümmler. Population structure of pseudomonas aeruginosa. *Proceedings of the National Academy of Science USA*, 104(19):8101–8106, May 2007.

[80] J. Zimakoff, N. Høiby, K. Rosendal, and J. P. Guilbert. Epidemiology of pseudomonas aeruginosa infection and the role of contamination of the environment in a cystic fibrosis clinic. *Journal of Hospital Infection*, 4(1):31–40, 1983.

[81] C. E. Wainwright, M. W. France, P. O'Rourke, S. Anuj, T. J. Kidd, M. D. Nissen, T. P. Sloots, C. Coulter, Z. Ristovski, M. Har-

greaves, B. R. Rose, C. Harbour, S. C. Bell, and K. P. Fennelly. Cough-generated aerosols of pseudomonas aeruginosa and other gram-negative bacteria from patients with cystic fibrosis. *Thorax*, 64(11):926–31, Nov 2009.

[82] L. D. Knibbs, G. R. Johnson, T. J. Kidd, J. Cheney, K. Grimwood, J. A. Kattenbelt, P. K. O'Rourke, K. A. Ramsay, P. D. Sly, C. E. Wainwright, M. E. Wood, L. Morawska, and S. C. Bell. Viability of pseudomonas aeruginosa in cough aerosols generated by persons with cystic fibrosis. *Thorax*, 69(8):740–5, Aug 2014.

[83] D. Williams, E. Benjamin, S. Haldenby, M. J. Walshaw, M. A. Brockurst, C. Winstanley, and S. Paterson. Divergent, coexisting pseudomonas aeruginosa lineages in chronic cystic fibrosis lung infections. *American Journal of Respiratory and Critical Care Medicine*, 191(7):775–785, 2015.

[84] M. R. O'Carroll, M. W. Syrmis, C. E. Wainwright, R. M. Greer, P. Mitchell, C. Coulter, T. P. Sloots, M. D. Nissen, and S. C. Bell. Clonal strains of pseudomonas aeruginosa in paediatric and adult cystic fibrosis units. *European Respiratory Journal*, 24(1):101–6, Jul 2004.

[85] L. Jelsbak, H. K. Johansen, A. Frost, R. Thøgersen, L. E. Thomsen, O. Ciofu, L. Yang, J. A. J. Haagensen, N. Høiby, and S. Molin. Molecular epidemiology and dynamics of pseudomonas aeruginosa populations in lungs of cystic fibrosis patients. *Infection and Immunity*, 75(5):2214–24, May 2007.

[86] J. L. Fothergill, M. J. Walshaw, and C. Winstanley. Transmissible strains of pseudomonas aeruginosa in cystic fibrosis lung infections. *European Respiratory Journal*, 40(1):227–238, Jul 2012.

[87] E. Gullberg, S. Cao, O. G. Berg, C. Ilbäck, L. Sandegren, D. Hughes, and D. I. Andersson. Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathogen*, 7(7):e1002158, Jul 2011.

[88] E. Kerem, S. Conway, S. Elborn, and H. Heijerman (Consensus Committee). Standards of care for patients with cystic fibrosis: a european consensus. *Journal of Cystic Fibrosis*, 4(1):7–26, Mar 2005.

[89] J. L. Burns, R. L. Gibson, S. McNamara, D. Yim, J. Emerson, M. Rosenfeld, P. Hiatt, K. McCoy, R. Castile, A. L. Smith, and B. W. Ramsey. Longitudinal assessment of pseudomonas aeruginosa in young children with cystic fibrosis. *Journal of Infectious DiseasesI*, 183(3):444–52, Feb 2001.

[90] H. K. Johansen, L. M. M. Sommer, R. L. Marvig, L. R. Jensen, M. Skov, T. Pressler, and S. Molin. What makes pseudomonas aeruginosa persist in the liungs of cf patients? *Pediatric Pulmonology*, 50(S41):S77–S107, Oct 2015.

[91] T. W. R. Lee, K. G. Brownlee, S. P. Conway, M. Denton, and J. M. Littlewood. Evaluation of a new definition for chronic pseudomonas aeruginosa infection in cystic fibrosis patients. *Journal of Cystic Fibrosis*, 2(1):29–34, Mar 2003.

[92] N. Hoiby, E. W. Flensborg, B. Beck, B. Friis, S. V. Jacobsen, and L. Jacobsen. Pseudomonas aeruginosa infection in cystic fibrosis. diagnostic and prognostic significance of pseudomonas aeruginosa precipitins determined by means of crossed immunoelectrophoresis. *Scandinavian Journal of Respiratory Diseases*, 58(2):65–79, Apr 1977.

[93] N. Jiricny, S. Molin, K. Foster, S. P. Diggle, P. D. Scanlan, M. Ghoul, H. K. Johansen, L. A. Santorelli, R. Popat, S. A. West, and A. S. Griffin. Loss of social behaviours in populations of pseudomonas aeruginosa infecting lungs of patients with cystic fibrosis. *PLoS One*, 9(1):e83124, 2014.

[94] O. Ciofu, B. Lee, M. Johannesson, N. O. Hermansen, P. Meyer, and N. Høiby (Scandinavian Cystic Fibrosis Study Consortium). Investigation of the algt operon sequence in mucoid and non-mucoid pseudomonas aeruginosa isolates from 115 scandinavian patients with cystic fibrosis and in 88 in vitro non-mucoid revertants. *Microbiology*, 154(Pt 1):103–13, Jan 2008.

[95] R. L. Marvig, L. M. Sommer, L. Jelsbak, S. Molin, and H. K. Johansen. Evolutionary insight from whole-genome sequencing of pseudomonas aeruginosa from cystic fibrosis patients. *Future Microbiology*, 10(4):599–611, 2015.

[96] J. C. Thøgersen, M. Mørup, S. Damkiær, S. Molin, and L. Jelsbak. Archetypal analysis of diverse pseudomonas aeruginosa transcriptomes reveals adaptation in cystic fibrosis airways. *BMC Bioinformatics*, 14:279, 2013.

[97] J. C. Thøgersen. *Metabolic adaptation of a human pathogen during chronic infections - a systems biology approach*. PhD thesis, Technical University of Denmark, Anker Engelunds vej 1, 2800 Lyngby, 2015.

[98] A. L. Goodman, B. Kulasekara, A. Rietsch, D. Boyd, R. S. Smith, and S. Lory. A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in pseudomonas aeruginosa. *Developmental Cell*, 7(5):745–54, Nov 2004.

[99] M. Fegan, P. Francis, A. C. Hayward, G. H. Davis, and J. A. Fuerst. Phenotypic conversion of pseudomonas aeruginosa in cystic fibrosis. *Journal of Clinical Microbiology*, 28(6):1143–6, Jun 1990.

[100] S. S. Pedersen, N. Høiby, F. Espersen, and C. Koch. Role of alginate in infection with mucoid pseudomonas aeruginosa in cystic fibrosis. *Thorax*, 47(1):6–13, Jan 1992.

[101] P. K. Singh, A. L. Schaefer, M. R. Parsek, T. O. Moninger, M. J. Welsh, and E. P. Greenberg. Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature*, 407(6805):762–4, Oct 2000.

[102] L. Yang, J. A. J. Haagensen, L. Jelsbak, H. K. Johansen, C. Sternberg, N. Høiby, and S. Molin. In situ growth rates and biofilm development of pseudomonas aeruginosa populations in chronic lung infections. *Journal of Bacteriology*, 190(8):2767–76, Apr 2008.

[103] M. C. Alanin, K. G. Nielsen, C. von Buchwald, M. Skov, K. Aanaes, N. Høiby, and H. K. Johansen. A longitudinal study of lung bacterial pathogens in patients with primary ciliary dyskinesia. *Clinical Microbiology and Infection*, Sep, ahead of print 2015.

[104] J. R. W. Govan and P. H. Brown. Evidence for transmission of pseudomonas cepacia by social contact in cystic fibrosis. *Lancet*, 342(8862):15–19, Jul 1993.

[105] B. Frederiksen, C. Koch, and N. Høiby. Changing epidemiology of pseudomonas aeruginosa infection in danish cystic fibrosis patients (1974-1995). *Pediatric Pulmonology*, 28(3):159–166, 1999.

[106] E. Mahenthiralingam, M. E. Campbell, and D. P. Speert. Nonmotility and phagocytic resistance of pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis. *Infection and Immunity*, 62(2):596–605, Feb 1994.

[107] L. R. Hoffman, E. Déziel, D. A. D'Argenio, F. Lépine, J. Emerson, S. McNamara, R. L. Gibson, B. W. Ramsey, and S. I. Miller. Selection for staphylococcus aureus small-colony variants due to growth in the presence of pseudomonas aeruginosa. *Proceedings of the National Academy of Science USA*, 103(52):19890–5, Dec 2006.

4 PRESENT INVESTIGATIONS

4.1 study 1:

Investigations of inter- and intra-clonal diversity of Pseudomonas aeruginosa populations in cystic fibrosis patients.

Sommer L. M., Marvig R. L., Luján A., Koza A., Pressler T., Molin S., and Johansen H. K. (2015), *Manuscript in preparation*

Investigations of inter- and intra-clonal diversity of *Pseudomonas aeruginosa* populations in cystic fibrosis patients

Lea Mette Sommer, [a]# Rasmus Lykke Marvig,[b] Adela Luján,[c] Anna Koza,[a] Tacjana Pressler,[d] Søren Molin,[a,e] Helle Krogh Johansen,[a,b]

The Technical University of Denmark, Center for Biosustainability, Hørsholm, Denmark[a]; Department of Clinical Microbiology, Rigshospitalet, Copenhagen, Denmark[b]; Exeter University, Cornwall, England[c]; CF Centre, Rigshospitalet, Copenhagen, Denmark[d]; Technical University of Denmark, Systems Biology, Lyngby, Denmark[e]

Running Head: Diversity of *P. aeruginosa* in CF airways

#Corresponding author
Lea M. Sommer, lemad@bio.dtu.dk

ABSTRACT

The primary cause of morbidity and mortality in cystic fibrosis (CF) patients is lung infection by *Pseudomonas aeruginosa*, a versatile opportunistic pathogen. Therefore much work has been done to understand the adaptation and evolution of *P. aeruginosa* to the CF lung. However, many of these studies have focused on longitudinally collected single isolates, and only few have included cross-sectional analyses of the whole *P. aeruginosa* populations in sputum samples. To date only few studies have used the approach of metagenomic analysis for the purpose of investigating *P. aeruginosa* populations in CF airways.

Here we combine the approach of analysing single longitudinally collected samples with a metagenomic analysis of single sputum samples, with the aim of investigating whether the genomic information from single isolates is sufficient to document evolutionary and adaptational patterns of the *P. aeruginosa* populations within CF patients' airways.

We analysed five metagenomes together with longitudinally collected single isolates from four CF patients. With this approach we were able to link the clone type and the SNP profile (or part of) of the single isolates to that of the metagenome(s) for each individual patient.

Based on our analysis we conclude that using comprehensive collections of longitudinal single isolates it is possible to rediscover the genotypes of the single isolates in the metagenomic samples. This indicates that the information gained by the genome sequencing of single isolates is sufficient for the research of adaptation and evolution of *P. aeruginosa* to the CF airways.

INTRODUCTION

Cystic fibrosis (CF) is a hereditary disease that causes malfunction of a chloride channel affecting the viscosity of the mucus on all muco-epithelial surfaces. Among other things, this results in impaired clearance of bacteria and other microorganisms from the airways with an associated increased risk of lung infections[1]. CF is the most common life-limiting genetic disorder in Caucasians, and lung infection with *Pseudomonas aeruginosa* is the primary cause of morbidity and mortality in CF patients[2,3]. In the clinic, antibiotic treatment of these infections is usually based on the assumption that the bacterial populations in CF airways are homogeneous. In accordance with this assumption, several studies of the adaptation of *P. aeruginosa* to the CF airway environment with regard to e.g. resistance development[4,5], metabolism[6], avoidance of the immune system[7], and transmission between niches in the airways of a patient[8] and between patients[9,10], have primarily been carried out based on investigations of single longitudinally stored bacterial isolates[11–13].

However, it was recently shown that long-term bacterial infections of CF airways cannot solely be described as a "dominant lineage" model, where the infecting clone type adapts in a linear fashion, and new variants with increased fitness quickly outcompete their less fit ancestors[14]. Because of the heterogeneous environment of the CF airways, it is more likely a "diverse-community" model that best describes the bacterial populations of the CF airways. This is an effect of adaptive radiation and the development of different subpopulations with a high degree of polymorphic mutations[11,14–16].

Thus, the question remains whether genomic information from single isolates collected longitudinally from the same patient is sufficient for the characterization of adaptive and evolutionary processes in *P. aeruginosa* populations in CF airways, and consequently also for the development of new diagnostic and therapeutic strategies. To answer this question, our approach has been to compare sequences from longitudinally collected single isolates with single metagenomes from four CF patients.

In this analysis, rediscovery in the metagenomes of the genome sequences derived from the single isolates has been considered evidence that the single isolates have proliferated in the population and thus are representative for the infecting population of the patient. If the single isolates are propagating lineages representing the adapting population, we would expect to find a clear correlation between their genomes and the metagenome(s). Moreover, we would only expect to find this correlation between isolates and metagenomes from the same patient, in contrast to cases of closely related bacterial populations established by patient-to-patient transmissions.

MATERIALS & METHODS

We included four CF patients followed at the CF-center at Rigshospitalet, Copenhagen, Denmark. The age of the patients ranged from 15 to 31 years and they were all recently diagnosed as chronically infected with *P. aeruginosa* (Copenhagen criteria[17]).

Longitudinally collected single isolates

Genome sequenced longitudinally collected single isolates from the patients are described in details in Marvig *et al.* (2014) [18]. The single isolates included in this study cover *P. aeruginosa* sampled from: endolaryngeal suction, sputum samples, sinus samples taken at endoscopic sinus surgery, swabs from the sinuses, and broncho alveolar lavage (BAL) samples.

Metagenomic samples

Sputum samples were collected at the CF clinic at Rigshospitalet and samples were processed a median of two days after expectoration (range: one to three days). During the lag-time between expectoration and processing the samples were stored at 4°C.

Processing of metagenomic samples

The samples were treated with ca. 1:1 (v/v) 10x diluted Sputasol (Oxoid, c/o Thermo Fisher Scientific, UK) with continuous vigorous shaking for 30 min. for homogenisation.

The samples were divided into two fractions, one was plated on PIA plates and incubated in 24-72h at 37°C depending on when colonies appeared and before single colonies could no longer be picked. The single colonies were then grown in 96 well microtitter plates with 150µl Luria Broth (LB) for 24-48h. 100µl 50% glycerol was added and the isolates were stored at -80°. The other fraction was directly subjected to DNA extraction (200-600µl Sputasol treated sample), or stored at -20°C until batch DNA purification could take place.

DNA extraction was carried out as in Lim *et al.* (2013)[19], with slight modifications: β-mercapoethanol was replaced by Sputasol treatment, centrifugation times were extended to 20 min at 3800x g, the volumes were adjusted to: 1.5ml autoclaved milliQ, 100-200µl DNase buffer and 3-6µl DNase (depending on pellet size), and 1.5 SE buffer, with the Powersoil® DNA isolation kit (MO BIO Laboratories, USA) used according to the manufacturer, for DNA purification.

Sequencing and analysis of metagenomic reads

Libraries were prepared in triplicates with Nexter XT Sample Preparation kit (Illumina Inc., USA) and pooled prior to sequencing on an Illumina MiSeq® bench-top sequencer with MiSeq reagent kit V2, 300 cycles (Illumina Inc., USA), resulting in 150bp paired end reads.

Initial analysis of the reads (also used for species identification) was carried out using Novoalign V2.07.18 (Novocraft Technologies[20]) for alignment to a library of human (GRCh37, ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens/Assembled_chromosomes/seq/hs_ref_GRCh37.p13_chr*mfa.gz), bacterial, archaeal, viral, and fungal (The NCBI database, downloaded: November 11th 2013) sequences. All sequences aligning to the human genome were discarded.

The analysis of the *P. aeruginosa* population was performed according to Marvig *et al.* (2014)[18]: Alignment of reads to PAO1 simultaneously removing all non-*P. aeruginosa* reads with SAM tools V0.1.7 (r50),[21] The Genome Analysis Toolkit (GATK) V1.0.5085 (compiled 2011/01/26 10:13:39),[22] and Bowtie2 V2.0.2[23].

SNP calling was carried out manually from via the .totalpileup files from Bowtie2 with the aid of information from the single isolates[18].
The rediscovery of SNPs from the single isolates in the metagenomes have been done on the assumption that if a SNP is present in more than 10% of the reads at the specific position of the SNP, this is regarded as a rediscovered SNP.
When looking at polymorphisms, only positions with a phred score >30 and with >=4 read coverage were considered. This was then compared to the coverage, where positions were only considered if the phred score was >30, to make the ratio presented in Figure 6.

Diversity measurement of metagenomic samples

Polymorphic positions were identified as described above. Because of the varying coverage of the different samples, the polymorphisms found were compared to the overall coverage of the PAO1 genome, and diversity was calculated as a ratio of polymorphisms and coverage. The reason, for using polymorphisms as a diversity measurement, is based on the assumption that the more positions a population diverge in, the more diverse is it likely to be.
To explain: if we have e.g. two or three subpopulations they will differ from each other in a number of positions creating more ambiguous base calls and thus a higher ratio of polymorphisms, than a single homogeneous population. It is important to keep in mind, that it is not possible to differentiate between two, three, or more different subpopulations based on this method. This is because of the possibility of a deep phylogenetic branching of two subpopulations and a shallow branching of three or more subpopulations.
However, the importance of this differentiation can be discussed, as it has always been difficult to determine diversity depending on, more or less subpopulations and higher or lower degrees of phylogenetic branching.

Diversity measurement by phylogenetic analysis of single isolates

Diversity is shown as the mean distance to the Line of Decent (LOD)[12]. LOD is the immediate line from the root (here based on PAO1 as out-group) to the latest sampled isolate (the red line in the phylogenetic trees in Figure 5), and the mean distance to LOD is the mean number of SNPs from this line to the remaining isolates.

Statistics

All statistics were done in R[24].

Ethics

The local ethics committee at the Capital Region of Denmark Region Hovedstaden approved the use of the samples: registration number H-4-2015-FSP. All patients

have given informed consent. In patients bellow 18 years of age, informed consent was obtained from their parents.

RESULTS

Patient information and *P. aeruginosa* infection patterns

Four CF patients were enrolled in this study (median age 24 years; range 15-31, at the time of metagenome sampling), from each of whom we have previously genome sequenced 9 to 27 longitudinally collected single *P. aeruginosa* isolates[18]. From the four patients we collected either one (n=3) or two (n=1) sputum samples for metagenomic analysis (Figure 1A).

Accordingly, sputum samples S1, S2, and S3 were sampled from patients P41M3, P99F4 and P92F3, and sputum samples S4a and S4b, separated by two weeks, were sampled from patient P82M3. The sputum samples used for the meta-genome sequencing were collected approximately one year after the most recently genome sequenced single isolate.

The time period between the most recently genome sequenced isolate and the metagenome is not critical, since the main question addressed here is whether or not the genotypes of the single isolates can be rediscovered in the metagenomic samples. If so, this would indicate that the respective single isolates are representative of a dominant propagating genotype(s) of the *P. aeruginosa* population of the patients.

Three of the four patients (P41M3, P92F3, and P82M3) have infection patterns that are characteristic for the majority of the *P. aeruginosa* infected CF patients at the Copenhagen CF Center at Rigshospitalet[25], with a single primary clone type in the entire collection period. One patient (P99F4) has a change in clone type, where one clone type is outcompeted by another (Figure 1A). Together the four patients display infection patterns representing approximately 90% of the Copenhagen CF children and young adults, as previously described[18,25].

All four patients in this study have been recently diagnosed as chronically infected with *P. aeruginosa* according to the Copenhagen definitions at the time of metagenome sampling[17].

Processing of sputum sample reads

The metagenome sequences were aligned to a database containing all bacterial, fungal, and viral genome sequences deposited at NCBI (see Materials and Methods). With a median of 96.11% of all bacterial reads *P. aeruginosa* was the dominating microbial species in the patients, corresponding to their clinical diagnosis as chronically infected with *P. aeruginosa*.

We further aligned reads from the sputum metagenomes to the *P. aeruginosa* PAO1 reference genome, as we have previously done for the single isolates[18]. In all cases, the metagenomic reads aligned to nearly all positions in the 6.3 Mbp reference genome, and on average 5.9 Mbp were covered by >3 reads. This high genomic coverage ensured that the presence or absence of polymorphisms in the metagenomes could be determined at the majority of genomic positions. On average, sequenced positions were covered by 10 to 31 reads (Supplementary Table S1) giving us the opportunity to identify subpopulations that are present in more than 10% of the population at the positions with the lowest coverage.

In order to compare the *P. aeruginosa* population structure and diversity as displayed by the single isolates and the compliance with the metagenomic read assemblies, we conducted a three step analysis: 1) Identification of the dominant clone type(s) in the sputum samples, 2) investigation of mutations in the genomes of the single isolates also identified in the metagenomes, i.e. rediscovery of SNPs in the metagenomes, and (3) comparison of diversity measurements of the populations represented by the single isolates and the metagenomes.

Identification of the dominant clone type(s)

To identify the *P. aeruginosa* clone types represented in the metagenomes, de novo assemblies of single isolates and metagenomes were compared, using dnadiff from MUMmer, with default parameters. That is, for each patient the clone types represented by the single isolates within a patient were compared with the metagenome(s) from the same patient.

For all four patients, the clone type of the most recently sampled single isolate corresponded to the clone type identified from the metagenome with less than 528 bp of difference (median 131 bp, range 91-527 bp). In contrast, when comparing the metagenomes with single isolates of other clone types they differed by more than 17,843 bp (median 17,844 bp, range 16,269-30,918 bp) (Supplementary Table S2). This shows that the most recent clone type identified by the genomes of the single isolates of each patient match the dominating clone type in the *P. aeruginosa* populations as identified in the sputum sample metagenomes.

Rediscovery of SNPs in the metagenomes

Previous investigations of genome evolution in the clonal lineages of *P. aeruginosa* strains from each of the four patients[18] led to identification of SNPs accumulating in the clonal populations present in the individual patients. If these SNPs are indeed present in actual propagating lineages of the *P. aeruginosa* population of these patients, they should also be present in the metagenome(s). When looking at all the SNPs identified in all the single isolates, it is expected that the ratio of rediscovery of SNPs between single isolates and metagenomes from the same patients should

exceed the ratio determined between single isolates and metagenomes of different patients. Further, this ratio should reach a value of one if all mutations found in the single isolates are also present in the metagenome.

With the exception of patient P99F4 and P92F3, who share the same clone type (DK26), the rediscovery of SNPs from isolates in metagenomes of the same patient was found to be significantly higher than between patients (Figure 2, $p<0.05$, Fisher's exact test with Holm correction). This supports the specific linkage between single isolates and the *P. aeruginosa* population as a whole, as hypothesised above.

In one case (S2 from P99F4), the ratio of the rediscovery of SNPs reached one, suggesting that all SNPs identified in the single isolates are present in more than 10% of the population as a whole. In all other cases the ratio was below one, which could be due to: 1) not all mutations being fixed in the population, i.e. they were lost during the time of sampling of the single isolates (harbouring the mutations) until sampling of the metagenome, or 2) some of the mutations being present in only a small fraction (<10%) of the population and therefore not sampled by the metagenomic reads. In the case of P92F3 the SNPs that are not rediscovered are only present in 11-22% of the single isolates, and thus could be explained by the first point stated above.

The metagenomes S4a and S4b from patient P82M3 illustrate both explanations: The first may explain the much lower ratio of rediscovery of SNPs in patient P82M3 compared to the other patients, because P82M3 has been shown to harbour hyper-mutators. Hyper-mutators are known to accumulate many unfavourable mutations[26], which are not expected to remain in the population, thus leading to a low ratio of rediscovery (assuming that the mutations are not hitch-hiking with more favourable mutations).

Concerning the second explanation regarding low coverage of the metagenomic samples, we found that in S4b 26% (54 of 461) of the SNPs from the single isolates of patient P82M3 were rediscovered (the latest metagenome), whereas only 12% (122 of 461) were identified in S4a (the first metagenome). This is contradictory since the mutations were previously identified in the single isolates and therefore must be present to some degree in S4a in order to be identified in S4b. This suggests that the subpopulation represented by the S4b metagenome is present below the limit of detection in the S4a metagenome sequences and is therefore not identified.

For the two patients: P99F4 and P92F3, the similar rediscovery ratios of SNPs between the metagenomes and the single isolates from the two patients can be explained by a co-infection of the same clone type, DK26.

This relationship was noted previously and seems to be the consequence of a patient-to-patient transmission event of the DK26 clone from P92F3 to P99F4[18], explaining the lack of differentiation between the *P. aeruginosa* populations.

However, despite this close relationship between the populations, Figure 3 shows

that it is possible to distinguish between the SNPs of the single isolates and the respective metagenomes.

We have identified SNPs in genome sequences of longitudinal single isolates, which seem to be specific and representative for the patient specific community, including cases of infections caused by patient-to-patient transmitted clones. A phylogenetic analysis of the single isolates and metagenomes of the hyper-mutator population of patient P82M3 underlines the patient specific relationship between metagenomes and single isolates.

Figure 4 shows that the metagenomes are placed within the phylogeny of the single isolates, confirming that the mutations identified in the single isolates are also identified in the respective metagenomes, and that the these single isolates therefore document the evolution and adaptation of the *P. aeruginosa* population.

Diversity of the *P. aeruginosa* populations

In the single isolates, the diversity of the *P. aeruginosa* populations was determined from the phylogenies as the mean distance to the Line of Decent (LOD), as previously described by Marvig (2013)[12] (Figure 5). For the metagenome-estimated diversity (Figure 6) we used the number of polymorphisms normalised to the number of positions covered in the PAO1 genome in order to correct for differences in coverage between the different metagenomes (see Materials and Methods for details). Because S4a and S4b (both from patient P82M3) are representative of the same population we chose to merge the samples to carry out the inter-patient comparison of diversity (Figure 6: S4, avg.).

In both the LOD calculations and the number of polymorphisms we find, not surprisingly, that the hyper-mutator population of patient P82M3 had the highest diversity and that patients with the shortest period of infection/colonisation (P92F3) as expected harboured the least diverse population. We calculated LODs of 34.89 and 1.33 mean distances to LOD for the two single isolate populations, and diversity ratios of 7.08E-5, and 4.20E-05 for the metagenome populations from the two patients P82M3 and P92F3, respectively. Thus, in both cases of diversity measurements (single isolates and metagenomes) we see a significant difference between the diversity of the *P. aeruginosa* populations of patient P82M3 and P92F3 ($p<0.05$, Fisher's Exact test with Holm correction) (Figure 5 and Figure 6).

When analysing further the single population of patient P82M3, the diversity calculations for the samples S4a and S4b illustrate that exhaustive sampling is essential, not only when using single isolates but also for metagenomic samples, in order to get the true picture of the population diversity. Because these two metagenomes represent a non-mutator and a hyper-mutator subpopulation,

respectively, they have significantly different diversity ratios (4.90E-05 and 9.25E-05, respectively, $p<0.05$ Fisher's Exact test with Holm correction).

In general, single isolates and metagenome analyses are dependent on exhaustive sampling and because of the possibility of temporal dominance by different subpopulations, as seen by the hyper-mutator population of P82M3, the metagenomic approach will also require multiple samples to reveal the complete profile of the *P. aerugonosa* population, especially in cases of high diversity, e.g. mutator populations.

DISCUSSION

Several investigations of long-term bacterial infections based on the analysis of longitudinally collected single isolates have been published in recent years[5,6,12,13,27,28], and some have included cross-sectional analyses of the population diversity at the genomic level[11,16,29,30].

One reason to question the validity of using single isolates to infer evolutionary dynamics of the entire population is the apparent heterogeneity of the *P. aeruginosa* population in the CF patients[11,15,16,29–32]. This heterogeneity may be the result of spatial heterogeneity of the CF airways and the confinement of different subpopulations to different niches[11,31]. Or, a skewed community composition, where one lineage is dominant where others are rare, in a form of frequency dependence[32].

Our results and previous findings by Marvig (2015)[13] do not confirm spatial isolation of different subpopulations in upper and lower airways (Figure 4) as has been suggested by Markussen (2014)[11] and Hansen (2012)[31]. In accordance with Ciofu (2013)[33] and Johansen (2012)[34] we find that the same genetic populations occupy both the upper and lower airways. Mixing of bacterial populations colonizing different airway compartments is also supported by other studies showing both genotypic and phenotypic overlap between samples from the upper and lower CF airways[33,35]. Population mixing is expected to result in more homogenous populations and might explain the low diversity in P99F4. However, we suspect that the low diversity found in this particular patient is better explained by a recent infection event.

Despite the evidence for population mixing, we and others[11,32] have observed relatively diverse populations of *P. aeruginosa* in the CF airways, more in line with a "Diverse Community" model as described by Lieberman et al. (2014)[14]. This model was used to describe the evolutionary and adaptational dynamics of *Burkholderia dolsa* in CF airways. It is further supported for *P. aeruginosa* by others e.g. Mowat (2011)[15], Feliziani (2014)[16], and Workentine (2013)[29], and has been confirmed for *Escherichia coli* in vitro in homogenous laboratory cultures[36,37].

In this study we have compared five meta-genomes obtained from four CF patient sputum samples with corresponding single, longitudinally collected, *P. aeruginosa* isolates, and a high degree of correlation was found. In addition, a highly diverse population was investigated by including two meta-genomes from a patient with previously identified hyper-mutator isolates. The meta-genomes were sampled from patients with the most common *P. aeruginosa* infection pattern at the Copenhagen CF Centre (Johansen H.K., unpublished), and they are therefore assumed to be representative for most of the CF patients.

One of the limitations of our study compared to that of Lieberman et al. (2014)[14] is the sequencing depth. This is especially true for the highly diverse population of P82M3, in which we were unable to identify subpopulations if present in less than 10.03% of the population (the lowest coverage is 9.97). Despite this, we were able to document the same diversification of populations as Lieberman (2014)[14] and others[16,38,39]. In addition, we were also able to determine that the different subpopulations comprising this diversity differ in frequency over time. Especially in the hyper-mutator population, we noticed that the bacterial population is dominated by different subpopulations at different time points (Table 1 and Figure 4).

If diversity is stably maintained you would expect that the main diversity of the metagenome is reflected by the bulk of the total SNPs identified in the single isolates from the specific patient. The *P. aeruginosa* population of patient P82M3 clearly shows fluctuation of the frequency of mutations comprising the main diversity at different time points. For the other three patients, especially P99F4 and P92F3, the diversity of the populations are maintained and most or all of the mutations found in all the single isolates are present in the metagenomes in a high degree.

Specifically the finding for the *P. aeruginosa* population of P82M3 is in agreement with results from Mowat (2011)[39], where they find that specific phenotypic subpopulations (haplotypes) have a higher chance of being resampled, and thus are more prevalent, than others at certain time points. Overall, they conclude that their results are consistent with either 1) continual emergence of new haplotypes, and/or 2) persistence of a diverse pool of haplotypes from which different haplotypes exhibit fluctuating relative abundance. Especially the second hypothesis is compatible with our results, from which we see the frequency of different mutations changing over a time period as short as two weeks (S4a and S4b). This is further consistent with the stable maintenance of most of the mutations from the early single isolates to the late sampled metagenomes.

It is important to stress that in our previous analysis in Marvig (2015)[13] we see a consistent pattern of mutations occurring over time during the infection. Genes such as gyrA/B, rpoB, mucA, algU, lasR, mexZ are continuously observed to accumulate

mutations independent of the study approach[11–13,16,27,31,40], indicating that evolutionary and adaptational patterns can be documented by investigations such as those applied here, including investigations of longitudinally collected single isolates.

In summary, our results show that it is possible to align population dynamic results obtained from genomic investigations of longitudinally collected single isolates of *P. aeruginosa* populations (metagenomes), as exemplified by the culture independent analysis of sputum samples.
The results of this study, taken together with similar results from other studies, suggest that using comprehensive collections of longitudinally collected single isolates in the research of adaptation and evolution of *P. aeruginosa* to the CF airways will yield useful results. However, considering the cost and ease of sequencing the use of metagenomic analyses for a deeper understanding of the population dynamics of *P. aeruginosa* evolution and adaptation is indeed encouraged.

Supplementary material:
Supplementary Table S1: Average coverage of bases that have been sequenced and aligned to the reference genome of PAO1.

Supplementary Table S2: Clone type comparisons between single isolates (Corresponding to the isolate numbers previously published in Study 2) and metagenomes from same patient. Next to the isolates of DK26 there is also written the patient from where it was sampled, since two patients in this study were colonised and infected with this clone type.

References

1. Gibson RL, Burns JL, Ramsey BW. Pathophysiology and management of pulmonary infections in cystic fibrosis. *J Respir Crit Care Med*. 2003;168(8):918-951.

2. Lyczak J, Cannon C, Pier G. Lung Infections Associated with Cystic Fibrosis. *Clin Microbiol Rev*. 2002;15(2). doi:10.1128/CMR.15.2.194.

3. O'Sullivan BP, Freedman SD. Cystic fibrosis. *Lancet*. 2009;373(9678):1891-1904. doi:10.1016/S0140-6736(09)60327-5.

4. Yonezawa M, Takahata M. DNA gyrase gyrA mutations in quinolone-resistant clinical isolates of Pseudomonas aeruginosa. *Antimicrob agents …*. 1995;39(9). doi:10.1128/AAC.39.9.1970.Updated.

5. Marvig RL, Søndergaard MSR, Damkiær S, et al. Mutations in 23S rRNA confer resistance against azithromycin in Pseudomonas aeruginosa. *Antimicrob Agents Chemother*. 2012;56(8):4519-4521. doi:10.1128/AAC.00630-12.

6. Marvig R, Damkiær S, Khademi S. Within-Host Evolution of Pseudomonas aeruginosa Reveals Adaptation toward Iron Acquisition from Hemoglobin. *MBio*. 2014. doi:10.1128/mBio.00966-14.Editor.

7. Feldman M, Bryan R, Rajan S, et al. Role of flagella in pathogenesis of Pseudomonas aeruginosa pulmonary infection. *Infect Immun*. 1998;66(1):43-51. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=107856&tool=pmcentrez&rendertype=abstract.

8. Mahenthiralingam E, Campbell ME, Speert DP. Nonmotility and Phagocytic Resistance of Pseudomonas aeruginosa Isolates from Chronically Colonized Patients with Cystic Fibrosis. *Infect Immun*. 1994;62(2):596-605. http://iai.asm.org/content/62/2/596.short. Accessed July 31, 2014.

9. McCallum SJ, Corkill J, Gallagher M, Ledson MJ, Hart C a, Walshaw MJ. Superinfection with a transmissible strain of Pseudomonas aeruginosa in adults with cystic fibrosis chronically colonised by P aeruginosa. *Lancet*. 2001;358(9281):558-560. http://www.ncbi.nlm.nih.gov/pubmed/11520530.

10. Jelsbak L, Johansen HK, Frost A-L, et al. Molecular epidemiology and dynamics of Pseudomonas aeruginosa populations in lungs of cystic fibrosis patients. *Infect Immun*. 2007;75(5):2214-2224. doi:10.1128/IAI.01282-06.

11. Markussen T, Marvig RL, Gómez-Lozano M, et al. Environmental Heterogeneity Drives Within-Host Diversification and Evolution of Pseudomonas aeruginosa. *MBio*. 2014;5(5):1-10. doi:10.1128/mBio.01592-14.Editor.

12. Marvig R, Johansen H, Molin S, Jelsbak L. Genome analysis of a transmissible lineage of Pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet*. 2013;9(9). doi:10.1371/journal.pgen.1003741.

13. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nat Genet*. 2015;47(1):57-64. doi:10.1038/ng.3148.

14. Lieberman TD, Flett KB, Yelin I, et al. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat Genet*. 2014;46(1):82-87. doi:10.1038/ng.2848.

15. Mowat E, Paterson S, Fothergill JL, et al. Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *Am J Respir Crit Care Med*. 2011;183(12):1674-1679. doi:10.1164/rccm.201009-1430OC.

16. Feliziani S, Marvig RL, Luján AM, et al. Coexistence and Within-Host Evolution of Diversified Lineages of Hypermutable Pseudomonas aeruginosa in Long-term Cystic Fibrosis Infections. *PLoS Genet*. 2014;10(10):e1004651. doi:10.1371/journal.pgen.1004651.

17. Johansen HK, Nørregaard L, Gøtzsche PC, Pressler T, Koch C, Høiby N. Antibody response to Pseudomonas aeruginosa in cystic fibrosis patients: a marker of therapeutic success?--A 30-year cohort study of survival in Danish CF patients after onset of chronic P. aeruginosa lung infection. *Pediatr Pulmonol*. 2004;37(5):427-432. doi:10.1002/ppul.10457.

18. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nat Genet*. 2015;47(1):57-64. doi:10.1038/ng.3148.

19. Lim YW, Schmieder R, Haynes M, et al. Metagenomics and metatranscriptomics: Windows on CF-associated viral and microbial communities. *J Cyst Fibros*. 2012;12(2):154-164. doi:10.1016/j.jcf.2012.07.009.

20. Krawitz P, Rödelsperger C, Jäger M, Jostins L, Bauer S, Robinson PN. Microindel detection in short-read sequence data. *Bioinformatics*. 2010;26(6):722-729. doi:10.1093/bioinformatics/btq027.

21. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352.

22. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498. doi:10.1038/ng.806.A.

23. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie2. *Nat Methods*. 2012;9(4):357-359. doi:10.1038/nmeth.1923.Fast.

24. R Core Team. R: A language and environment for statistical computing. 2013. http://www.r-project.org/.

25. Johansen HK, Madsen LM, Marvig RL, Pressler T, Molin S. Rethinking Pseudomonas aeruginosa (PA) lung infection: using molecular microbiology rather than culture and antibodies. *J Cyst Fibros*. 2014;13(Suppl. 2):S33.

523 26. Arjan J a, Visser M, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. Diminishing
524 returns from mutation supply rate in asexual populations. *Science*.
525 1999;283(5400):404-406. doi:10.1126/science.283.5400.404.

526 27. Yang L, Jelsbak L, Marvig RL, et al. Evolutionary dynamics of bacteria in a
527 human host environment. *Proc Natl Acad Sci U S A*. 2011;108(18):7481-7486.
528 doi:10.1073/pnas.1018249108.

529 28. Bragonzi A, Paroni M, Nonis A, et al. Pseudomonas aeruginosa
530 microevolution during cystic fibrosis lung infection establishes clones with adapted
531 virulence. *Am J Respir Crit Care Med*. 2009;180(2):138-145.
532 doi:10.1164/rccm.200812-1943OC.

533 29. Workentine ML, Sibley CD, Glezerson B, et al. Phenotypic Heterogeneity of
534 Pseudomonas aeruginosa Populations in a Cystic Fibrosis Patient. *PLoS One*.
535 2013;8(4):1-10. doi:10.1371/journal.pone.0060225.

536 30. Darch SE, McNally A, Harrison F, et al. Recombination is a key driver of
537 genomic and phenotypic diversity in a Pseudomonas aeruginosa population during
538 cystic fibrosis infection. *Sci Rep*. 2015;5. http://dx.doi.org/10.1038/srep07649.

539 31. Hansen SK, Rau MH, Johansen HK, et al. Evolution and diversification of
540 Pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have
541 implications for chronic lung infection. *ISME J*. 2012;6(1):31-45.
542 doi:10.1038/ismej.2011.83.

543 32. Williams D, Evans B, Haldenby S, et al. Divergent, coexisting Pseudomonas
544 aerugnosa lineages in chronic cystic fibrosis lung infections. *Am J Respir Crit Care*
545 *Med*. 2015;191(7):775-785. doi:10.1164/rccm.201409-1646OC.

546 33. Ciofu O, Johansen HK, Aanæs K, et al. P. aeruginosa in the paranasal sinuses
547 and transplanted lungs have similar adaptive mutations as isolates from chronically
548 infected CF lungs. *J Cyst Fibros*. 2013;12(6):729-736. doi:10.1016/j.jcf.2013.02.004.

549 34. Johansen HK, Aanaes K, Pressler T, et al. Colonisation and infection of the
550 paranasal sinuses in cystic fibrosis patients is accompanied by a reduced PMN
551 response. *J Cyst Fibros*. 2012;11(6):525-531. doi:10.1016/j.jcf.2012.04.011.

552 35. Mainz JG, Naehrlich L, Schien M, et al. Concordant genotype of upper and
553 lower airways P aeruginosa and S aureus isolates in cystic fibrosis. *Thorax*.
554 2009;64(6):535-540. doi:10.1136/thx.2008.104711.

555 36. Elena S, Lenski R. Long-term experimental evolution in Escherichia coli. VII.
556 Mechanisms maintaining genetic variability within populations. *Evolution (N Y)*.
557 1997;51(4):1058-1067. http://www.jstor.org/stable/2411035. Accessed January 8,
558 2015.

559 37. Kinnersley MA, Holben WE, Rosenzweig F. E Unibus Plurum: genomic
560 analysis of an experimentally evolved polymorphism in Escherichia coli. *PLoS Genet*.
561 2009;5(11):e1000713. doi:10.1371/journal.pgen.1000713.

562 38. Workentine ML, Sibley CD, Glezerson B, et al. Phenotypic heterogeneity of
563 Pseudomonas aeruginosa populations in a cystic fibrosis patient. *PLoS One*.
564 2013;8(4):e60225. doi:10.1371/journal.pone.0060225.

565 39. Mowat E, Paterson S, Fothergill JL, et al. Pseudomonas aeruginosa population
566 diversity and turnover in cystic fibrosis chronic infections. *Am J Respir Crit Care Med*.
567 2011;183(12):1674-1679. doi:10.1164/rccm.201009-1430OC.

568 40. Smith EE, Buckley DG, Wu Z, et al. Genetic adaptation by Pseudomonas
569 aeruginosa to the airways of cystic fibrosis patients. *Notes*. 2006;103(22).

570

FIGURES AND TABLES

Table 1: Examples of rediscovered SNPs in S4a and S4b. Examples of mutations not found in S4a but in S4b.

							Total number of reads on position		% of reads that equals the mutation	
Position	Ref.	Qry.	Mutation	Type	PA number	Gene name	S4a	S4b	S4a	S4b
5671	C	T	C1397T	Missense	PA0004	gyrB	15	13	7	100
2453983	G	A	G317A	Missense	PA2231	pslA	17	17	24	0
2640133	A	G	T127C	Missense	PA2386	pvdA	15	12	0	100
2926243	C	G	G190C	Missense	PA2586	gacA	5	11	0	100
3970113	G	A	G58S	Missense	PA3545	algG	15	11	87	0
3971271	G	A	G444S	Missense	PA3545	algG	19	13	0	100
5551035	A	C	T647G	Missense	PA4946	mutL	11	8	9	100
5677066	T	C	A793G	Missense	PA5040	pilQ	12	17	17	100
6028514	G	A	G973A	Missense	PA5361	phoR	5	12	100	0

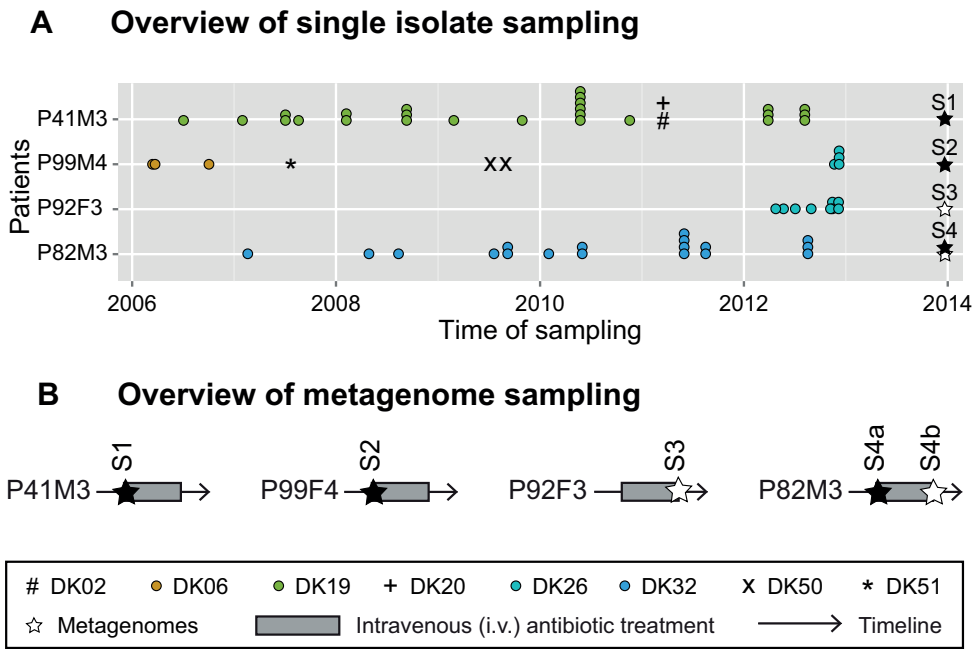
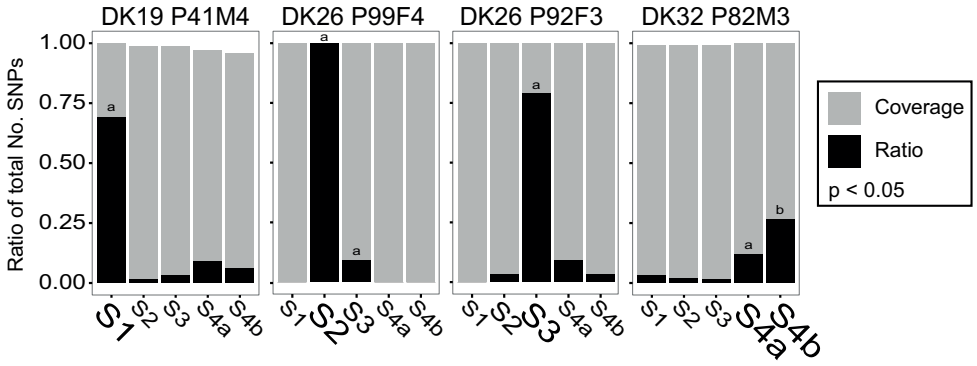
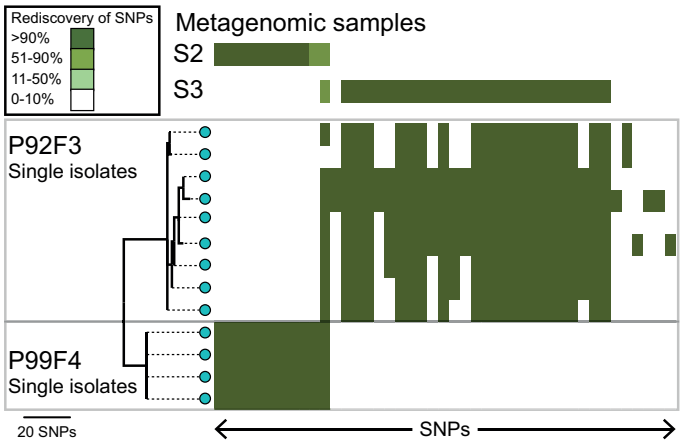


Figure 1, Overview. A) Overview of single isolate sampling from four CF patients, clone types that are considered to be transient (found in 1-2 time points) are marked with #, +, x, or *, whereas clones considered to be persistent in the patient is marked by coloured circles. Metagenomes are marked with stars, black if sampled before i.v. antibiotic treatment, and white if sampled after i.v. antibiotic treatment. **B)** Overview of metagenome sampling from patients, in correlation with two-week i.v. antibiotic treatment.



581 **Figure 2, SNP rediscovery in metagenomes.** Above each subfigure it is indicated
582 which single isolates' SNPs that have been sought rediscovered in the metagenomes
583 (clone type and patient). The grey bars indicate the ratio of SNP positions that were
584 sequenced in the metagenomes and the black bars indicate the ratio of the
585 rediscovered reads to the sequenced positions. The metagenomes belonging to the
586 same patient as the single isolates they are compared to is indicated with a larger
587 font. NOTE, S2 and S3 are from different patients but the same clone type. P<0.05,
588 Fisher's exact test with Holm correction, significant differences are indicated by "a"
589 and/or "b".



590 **Figure 3, Patient specific correlation of rediscovered SNPs within a single**
591 **clone type.** A comparison of the SNPs found in the single isolates of the patients
592 P99F4 and P92F3 as well as their respective metagenomes, S2 and S3. For the single
593 isolates a dark green colour indicates the presence of SNPs and white the absence.

For the metagenomes the green colours (dark, medium, light) indicates that >90%, 51-90%, and 11-50% of the reads covering the position of the SNP confirms the mutation, and if the SNP is confirmed with <=10% the SNP is not considered to be found in the metagenomes, indicated by white.

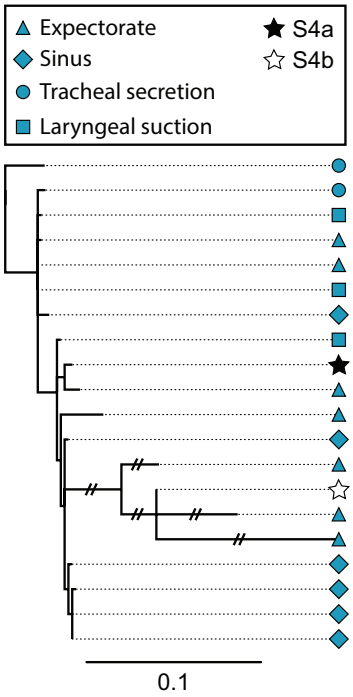
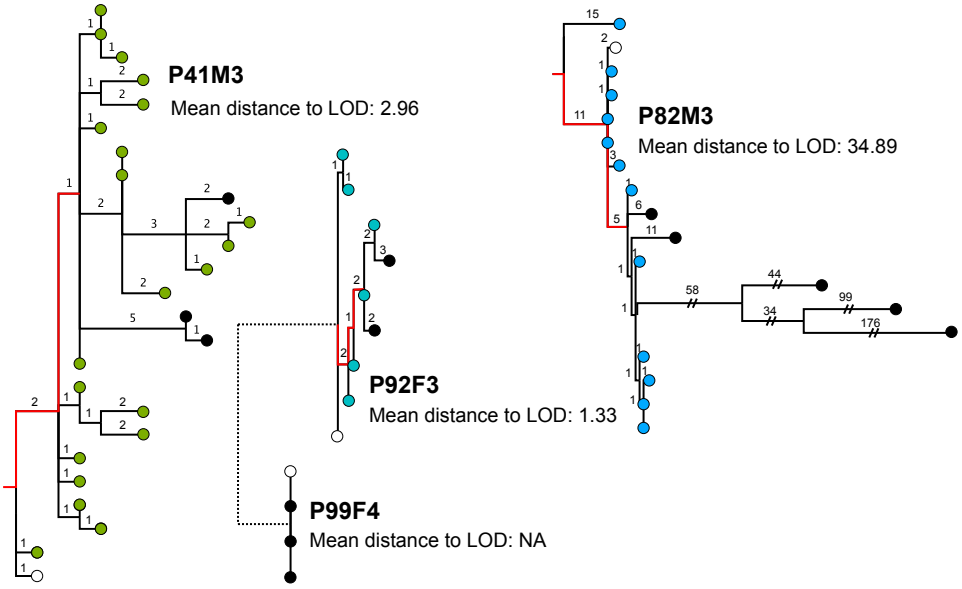
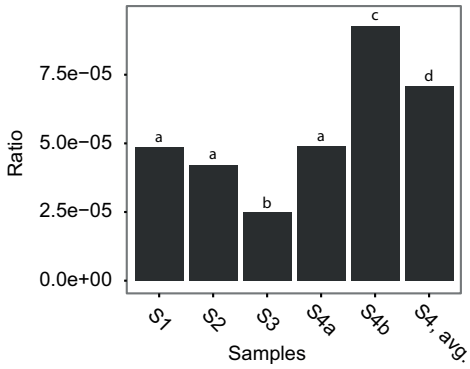


Figure 4: DK32 P82M3 maximum likelihood phylogeny including metagenomes. Blue shapes indicate single isolates sampled with different methods and from different locations (see legend) and stars indicate metagenomes. The scale bar indicates 0.1 likelihood of mutation.



603

604 **Figure 5, Mean distance to Line of Decent (LOD).** Maximum parsimonious
605 phylogenetic trees for all clone types identified in the metagenomes as being the
606 latest. The red line in the trees indicates the LOD wherefrom SNPs (numbers on
607 branches) have been counted. The LOD is set from the root (created for P41M3 and
608 P82M3 using PAO1 as out-group) to the divergence of latest sampled isolates. White
609 circles indicate the earliest sampled isolates and black circles indicates the latest
610 sampled isolates from each patient, the coloured circles are comparable to the
611 colours of Figure 1. For each patient a mean distance to LOD is indicated below the
612 patient name to the right of the corresponding tree. The LOD of P82M3 is
613 significantly different from the other patients, $p<0.01$, Fisher's Exact test with Holm
614 correction.



615 **Figure 6, Polymorphic positions in the metagenomes.** Because of the differing
616 coverage of the PAO1 reference genome, the number of polymorphisms is shown as
617 a ratio of polymorphisms and the coverage of each metagenome to PAO1. S4, avg. is
618 the average of S4a and S4b. $p<0.05$ Fisher's exact test with Holm correction.

Study 1 Supplementary:

Supplementary Table S1.
Average coverage of bases that have been sequenced and alligned to the refference genome of PAO1.

Sample	Total reads	Total bases of PAO1	Total bases sequenced of PAO1	Coverage of positions sequenced	Resolution (%)*
S1	159590945	6264404	6060807	26.33	3.80
S2	186456550	6264404	6084424	30.64	3.26
S3	123114192	6264404	6075272	20.26	4.93
S4a	59893402	6264404	6009803	9.97	10.03
S4b	68510406	6264404	6013647	11.39	8.78

*That is, if a subpopulation isto be detected by the sequencing it needs to be present with more than X% of the population.

Supplementary Table S2.
Clone type comparisons between single isolates (Corresponding to the isolate numbers previously published in Study 2) and metagenomes from same patient. Next to the isolates of DK26 there is also written the patient from where it was sampled, since two patients in this study were colonised and infected with this clone type.

Patient	REF Isolate	Clone	QRY Metagenome	Clone (presumed)	total bases (ref)	total bases (qry)	unaligned bases (ref)	%	unaligned bases (qry)	%	total GSNPs
P41M3	156	DK19	LM_4833	DK19	6559603	5992396	570509	8.70%	3686	0.06%	91
	382	DK26 (P99F4)	LM_4833	DK19	7002729	5992396	1108299	15.83%	93554	1.56%	30918
P99F4	390	DK06	LM_5085	DK26 (P99F4)	6676808	6002328	783036	11.73%	102156	1.70%	17339
	382	DK26 (P99F4)	LM_5085	DK26 (P99F4)	7002729	6002328	1010190	14.43%	2290	0.04%	131
	388	DK51	LM_5085	DK26 (P99F4)	6311260	6002328	361076	5.72%	44807	0.75%	17844
	386	DK50	LM_5085	DK26 (P99F4)	6261092	6002328	345715	5.52%	78021	1.30%	27331
P92F3	272	DK26 (P92F3)	LM_4711	DK26 (P92F3)	6666214	5993232	681864	10.23%	7738	0.13%	111
	156	DK19	LM_4711	DK26 (P92F3)	6559603	5993232	672137	10.25%	105926	1.77%	30862
P82M3	224	DK32	LM_4866	DK32	6371360	5606193	797887	12.50%	6501	0.12%	407
	224	DK32	LM_5137	DK32	6381360	5829566	571810	8.96%	5368	0.09%	527
	272	DK26 (P92F3)	LM_4866	DK32	6666214	5606193	1132616	16.99%	56614	1.01%	16269
	272	DK26 (P92F3)	LM_5137	DK32	6666214	5829566	909380	13.64%	58946	1.01%	17043

4.2 study 2:

Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis.

Marvig R. L., **Sommer L. M.**, Molin S., and Johansen H. K. (2015), Nature Genetics, **47**: p. 57-64. doi:10.1038/ng.3148

Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis

Rasmus Lykke Marvig[1,2], Lea Mette Sommer[2,3], Søren Molin[2,3] & Helle Krogh Johansen[1,3]

Little is known about how within-host evolution compares between genotypically different strains of the same pathogenic species. We sequenced the whole genomes of 474 longitudinally collected clinical isolates of *Pseudomonas aeruginosa* sampled from 34 children and young individuals with cystic fibrosis. Our analysis of 36 *P. aeruginosa* lineages identified convergent molecular evolution in 52 genes. This list of genes suggests a role in host adaptation for remodeling of regulatory networks and central metabolism, acquisition of antibiotic resistance and loss of extracellular virulence factors. Furthermore, we find an ordered succession of mutations in key regulatory networks. Accordingly, mutations in downstream transcriptional regulators were contingent upon mutations in upstream regulators, suggesting that remodeling of regulatory networks might be important in adaptation. The characterization of genes involved in host adaptation may help in predicting bacterial evolution in patients with cystic fibrosis and in the design of future intervention strategies.

A molecular understanding of how bacterial pathogens evolve during infection in their human hosts is important for the ability to treat infections. Advances in genome sequencing have made it possible to follow the evolution of bacteria by sequencing the genomes of the same strain over shorter or longer time periods[1–12]. Genome sequencing of bacterial pathogens has shown how the pathogenicity of bacterial clones can evolve via mutational changes in preexisting genes, a mechanism also known as pathogenicity or pathoadaptive mutation[13]. Although many such studies have provided insight into the genomic evolution of single clonal lineages (or clonal complexes) of human pathogens, little is known about how evolutionary paths compare between large numbers of genotypically different strains of the same species. For example, it remains unclear (i) to what extent natural selection restricts lineages with different genetic backgrounds to common pathways leading to adaptive phenotypes and (ii) in which way evolutionary outcomes become intertwined over time such that future alternatives may be contingent on the previous history of an evolving population.

To address these questions about evolutionary convergence and the role of historical contingency, we investigated the molecular evolution of distinct clonal lineages of *P. aeruginosa* from initial invasion of cystic fibrosis airways onward, as the lineages genetically adapted to a human host after transition from their environmental habitat. Our study was based on a large and diverse collection of 474 longitudinally collected isolates of *P. aeruginosa* sampled from the airways of patients with cystic fibrosis, in which *P. aeruginosa* was the predominant pathogen associated with increased morbidity and mortality[14].

RESULTS

Clinical collection of genome-sequenced bacterial isolates

Patients with cystic fibrosis attending the Copenhagen Cystic Fibrosis Center are closely followed, and sputum samples are cultured monthly to identify cystic fibrosis–associated bacterial pathogens at an early stage[15]. As a consequence, sputum samples are tested every month for the presence of *P. aeruginosa*, and antibiotic chemotherapy is initiated whenever *P. aeruginosa* is cultured (Online Methods). Furthermore, *P. aeruginosa* isolates are freeze stored for further characterization (**Fig. 1**). In this study, we have sequenced the whole genomes of a total of 474 isolates of *P. aeruginosa* sampled from the airways of 34 children and young individuals with cystic fibrosis (median age at first sequenced *P. aeruginosa* isolate = 8.8 years, range = 1.4–26.3 years). To obtain a bacterial collection that would give insight into the longitudinal progression of the early phase of infection, we chose to sequence the genomes of initial and subsequent isolates of *P. aeruginosa* from each of the subjects. On average, we sequenced the genomes of 12.9 isolates (range = 3–28 isolates) from each subject (**Fig. 2** and **Supplementary Fig. 1**), and the average time span for the sequenced isolates of *P. aeruginosa* from each subject was 4.8 years (range = 1–10 years) (**Fig. 2** and **Supplementary Fig. 2**).

Phylogenetically distinct clone types of *P. aeruginosa*

We compared the genomes of the 474 sequenced isolates and found that they grouped into 53 genetically distinct clone types (clonal complexes[16]) (**Fig. 1b**). Genomes for isolates of the same clone type differed on average by 122 SNPs (median = 9 SNPs, range = 0–1,333 SNPs), whereas genomes from different clone types differed by >10,000 SNPs (**Fig. 1c**).

Common clone types and transmission

Children with cystic fibrosis are anticipated to acquire their first *P. aeruginosa* colonization from unique environmental clone types of *P. aeruginosa* that are naive to the airways of the human host[17]. In agreement with this, we found 43 of the 53 clone types in only a single

[1]Department of Clinical Microbiology, Rigshospitalet, Copenhagen, Denmark. [2]Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark. [3]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. Correspondence should be addressed to R.L.M. (rmarvig@gmail.com).

Received 24 February; accepted 27 October; published online 17 November 2014; doi:10.1038/ng.3148

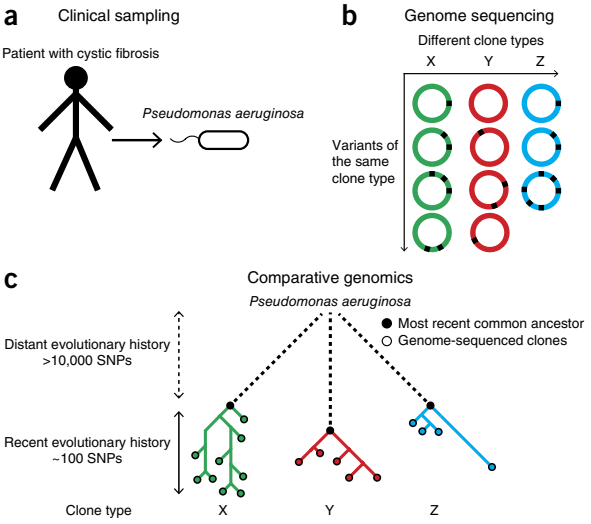
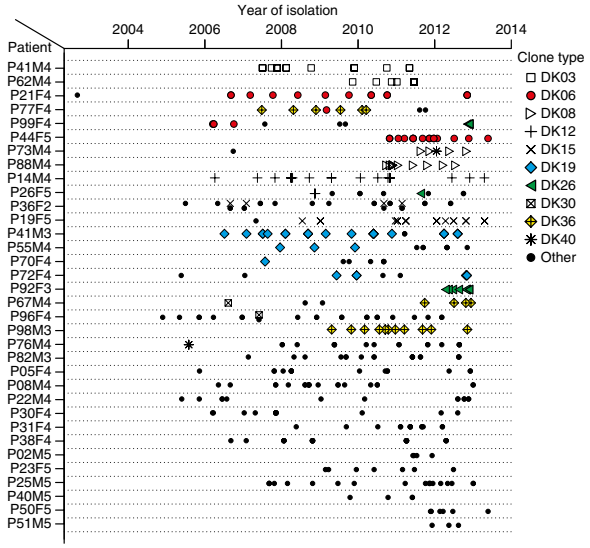


Figure 1 Overview of the present investigations. (**a**) To study within-patient evolution, we collected 474 isolates of *P. aeruginosa* from the airways of 34 patients with cystic fibrosis. (**b**) Genome sequencing identified 53 different clone types among the 474 isolates of *P. aeruginosa*. (**c**) Considerable genetic diversity existed between isolates of different clone types (>10,000 SNPs), whereas isolates from the same clone type differed by relatively few SNPs. On the basis of intraclonal genome comparisons, we reconstructed the phylogenetic history of each of the clonal complexes to gain insight into its recent evolutionary history, which most likely contains recent adaptive events that have occurred in the human host environment.

subject. However, we also identified ten clone types that were found in multiple subjects (**Fig. 2**).

The occurrence of the same clone type in several patients could be due to (i) direct patient-to-patient transmission in the outpatient clinic or in the ward, (ii) indirect transmission via instruments in the clinic or exposure to the same environmental reservoirs, or (iii) the existence of common environmental clone types[16,18]. When we measured the genetic distances between isolates of the same clone type sampled from different patients, we found that, in most cases (n = 19), at least 50 SNPs separated isolates from different patients (**Fig. 3a**). Taking into account that the within-patient mutation rate of *P. aeruginosa* is around 2.6 SNPs/year[1], we consider it unlikely that the presence of the same clone type in such cases is due to recent and/or direct patient-to-patient transmission. Instead, the presence of genetically distant isolates of the same clone type in these patients might be the result of transmission via patients not included in this study or the result of independent acquisitions of the clone type from the environment.

In a few cases (n = 5), only a few SNPs (range = 0–29 SNPs) differentiated clonal isolates from different patients (**Fig. 3a**), suggesting the occurrence of recent transmission events between patients, including P36F2 and P19F5 infected by clone type DK15; P21F4, P77F4 and P99F4 infected by clone type DK06; and P99F4 and P92F3 infected by clone type DK26. To investigate whether there were any epidemiological data to support recent transmission between these individuals, we retrieved information on the patients' visits to the hospital. In all cases of suggested patient-to-patient transmission, there were temporal overlaps in the respective patients' visits to the hospital



during the period before the detection of transmission (**Fig. 3b**). We presume that the clone types were transmitted from the older patients, in whom they were first identified, to the younger patients, in whom they were identified later. However, one should note that within-host population heterogeneity might confound such conclusions about the direction and source of transmission[19].

Identification of recent mutational events in clone types

It has previously been shown that *P. aeruginosa* genetically adapts to the human host environment[1–4]. Accordingly, each of the 53 clone types might have genetically adapted to human airways upon transition from the environment. To detect such recent mutational events in each clonal type, we compared the genomes of isolates from the same clone type to reconstruct the evolutionary history of each of the clonal lineages (that is, to identify mutations that have accumulated since the most recent common ancestor (MRCA)) (**Fig. 1c**). Note that such genetic comparisons were only possible for 36 of the 53 clone types, as 17 of the clone types were only represented by a single isolate.

The mutations in each of the lineages accumulated in a highly parsimonious fashion (average parsimonious consistency of 0.94; **Supplementary Table 1**), reflecting unidirectional and clonal evolution of the lineages since the MRCA. Thus, using a maximum-parsimony phylogenetic model, we were able to make inferences about the succession of mutations and the relationship among *P. aeruginosa* clones.

As a measure of the within-patient diversity of clonal populations, we counted the number of SNPs each isolate had accumulated since the MRCA of clonal isolates from the same patient. The median genetic distance to the MRCA of clonal isolates from the same patient was 8 SNPs, but 55 isolates belonging to 7 different clone types diverged from the MRCA by >50 SNPs (**Supplementary Fig. 3**).

Figure 2 Overview of the 474 genome-sequenced *P. aeruginosa* isolates. Each symbol represents a genome-sequenced isolate of *P. aeruginosa*. Axes indicate the patient from whom the isolate was derived and the time of sampling for each of the isolates. Specific symbols are indicated for isolates belonging to clone types (n = 10) that have been sampled from multiple patients. Colors are used where the same clone type was found in more than two patients (see **Fig. 3** for further details).

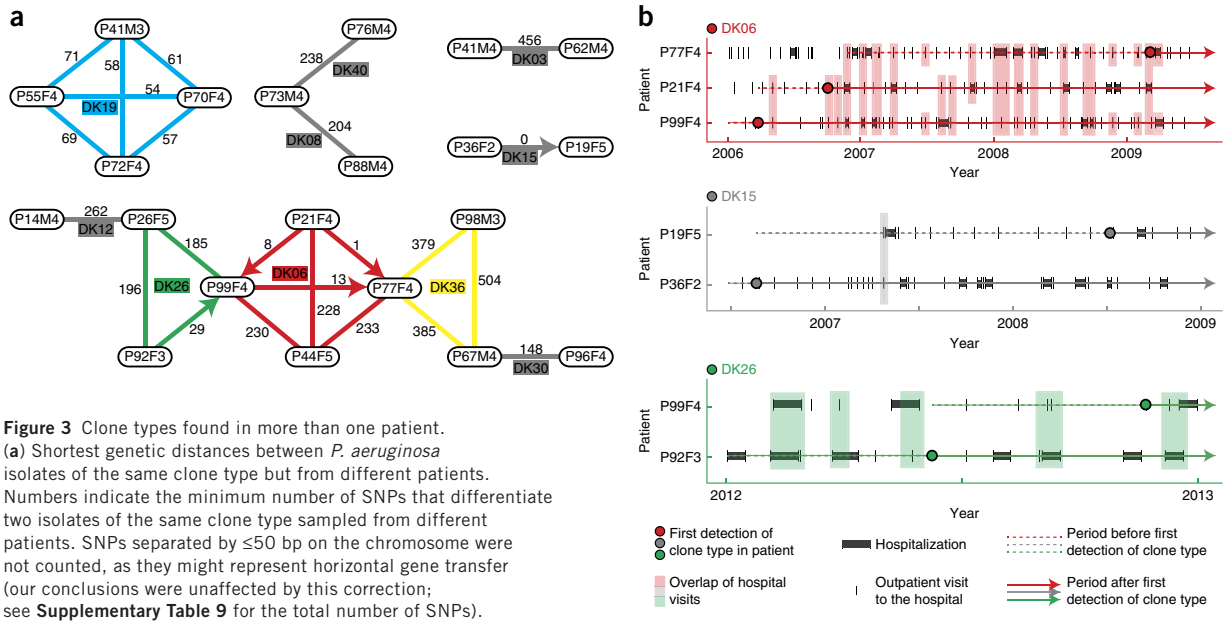


Figure 3 Clone types found in more than one patient. (**a**) Shortest genetic distances between *P. aeruginosa* isolates of the same clone type but from different patients. Numbers indicate the minimum number of SNPs that differentiate two isolates of the same clone type sampled from different patients. SNPs separated by ≤50 bp on the chromosome were not counted, as they might represent horizontal gene transfer (our conclusions were unaffected by this correction; see **Supplementary Table 9** for the total number of SNPs). Colors are used where isolates of the same clone type were found in more than two patients. Lines with arrows show the direction of patient-to-patient transmission as suggested by epidemiological data. (**b**) Overview of the patients' visits to the hospital (Copenhagen Cystic Fibrosis Center, Rigshospitalet). The phylogenetic data presented in **a** suggest that clone type DK15 was transmitted between patients P36F2 and P19F5, that clone type DK06 was transmitted between patients P21F4, P77F4 and P99F4, and that clone type DK26 was transmitted between patients P99F4 and P92F3. Vertical black lines show information about the patients' visits to the hospital in the period before detection of transmission. In all cases of suggested patient-to-patient transmission, the respective patients' visits to the hospital overlap temporally.

The large within-patient clone diversity in six of the seven lineages (DK01, DK12, DK15, DK32, DK36 and DK53) was due to an increased frequency of transition substitutions (note, in DK12, only part of the high degree of variation was due to an excess of transition substitutions; see the explanation below and **Supplementary Fig. 4**). This finding suggests hypermutation caused by defects in the DNA mismatch repair (MMR) genes[1,20], and, in agreement with this, we found that the six lineages were the only ones to have accumulated nonsynonymous mutations in one of the DNA MMR genes *mutS* ($n = 2$) and *mutL* ($n = 3$) or in both genes ($n = 1$). Nonetheless, no mutational signatures of hypermutation or of horizontal gene transfer were evident to explain the large within-patient variation of DK40 isolates from patient P73M4 (**Supplementary Fig. 5**). Similarly, part of the genetic variation among DK12 isolates from patient P44F5 could not be explained by hypermutation, as the particular isolates did not carry mutations in MMR genes nor exhibited mutational skews in transition-to-transversion ratios (**Supplementary Fig. 4**). Instead, we speculate that, in these two patients, the clonal variation was generated before infection and that the presence of genetically distant isolates of the same clone type is the result of multiple independent acquisitions of the same clone type from the environment or of a single acquisition of a diverse population.

Parallel evolution of genes involved in host adaptation

In total, we identified 12,324 mutations (9,517 SNPs and 2,807 small insertions and deletions (microindels)) that accumulated in the recent evolutionary history of the 36 clone types for which we had sampled multiple isolates (**Supplementary Tables 2** and **3**). We expect all clonal lineages to accumulate adaptive mutations in response to the environment of the human hosts in which they propagate. We therefore anticipate that parallelism might exist in the adaptive genetic routes of the different lineages. To identify such convergent evolution, we sought to identify genes that were mutated in parallel in multiple clone types. On the basis of the number of nonsynonymous mutations that accumulated in each lineage, we estimated the expected number of genes having nonsynonymous mutations in $\geq x$ clone types, considering a scenario in which mutations were randomly introduced by genetic drift (Online Methods and **Supplementary Table 4**). As larger genes are more likely to be randomly mutated, the criterion for which genes we would identify as subject to convergent evolution took into account the lengths of genes. Accordingly, we identified a total of 52 genes that were more frequently mutated than what would be expected under genetic drift (**Fig. 4** and **Supplementary Table 5**). Our findings suggest that parallel nonsynonymous mutations in these 52 genes are the result of positive selection for mutations in genes undergoing adaptive evolution. Therefore, we refer to these as candidate pathoadaptive genes in which mutations optimize pathogen fitness[1], although the presence of neutral mutational hotspots might contribute to the high mutation number in some genes.

We also note that, despite clinical tests for the presence of *P. aeruginosa* in the patients on a monthly basis, we cannot formally exclude the possibility that lineages might have acquired adaptive mutations that were already in the ancestor of the sampled isolates. Nonetheless, this possibility did not seem to confound our analyses, and, for example, no insertions or deletions were found to be fixed a priori in any of the lineages in the *mucA*, *lasR* and *rpoN* genes in which loss-of-function mutations are typical markers of cystic fibrosis–associated lung infections[2,14].

Function of pathoadaptive genes

We grouped the 52 candidate pathoadaptive genes according to their PseudoCap functions[21] (**Fig. 5a** and **Supplementary Table 5**),

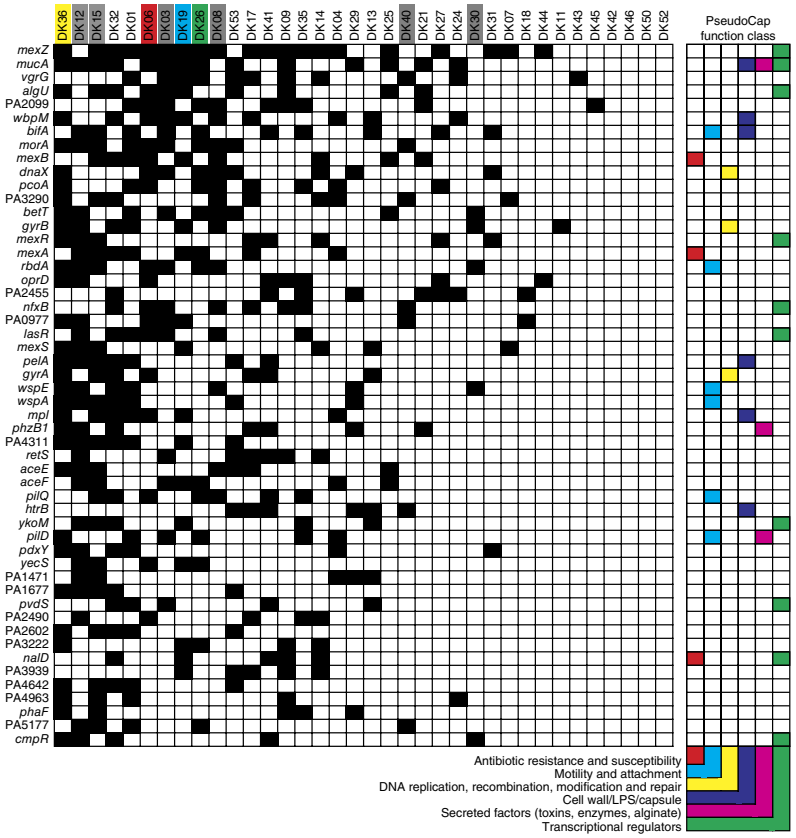


Figure 4 Pathoadaptive genes (n = 52). Genes identified on the basis of parallel evolution to be involved in host adaptation. Clone type names are highlighted with shaded boxes if the clone type was found in more than two patients (**Figs. 2** and **3**). The black squares in the large matrix (left) denote whether the genes underwent nonsynonymous mutation in the recent evolutionary history of the respective clone type. The colored squares in the small matrix (right) denote the PseudoCap[21] functions of the genes. See **Supplementary Table 5** for detailed information about the genes and number of mutations.

finding an over-representation of the classes 'antibiotic resistance and susceptibility', 'motility and attachment', 'DNA replication, modification and repair', 'cell wall/LPS/capsule', 'secreted factors (toxins, enzymes, alginate)' and 'transcriptional regulators' (**Supplementary Table 6**). These results are in line with previous studies that have reported that *P. aeruginosa* host adaptation is mediated by loss of motility, acquisition of antibiotic resistance, remodeling of regulatory networks, loss of extracellular virulence factors and modification of the cell envelope[1,22–29]. Furthermore, we anticipate that mutations, in general, are associated with loss of gene function, as 45 of the 52 genes were targeted by frameshift mutations (with the exceptions of *gyrA*, *gyrB*, PA1471, *pcoA*, *mexS*, *vgrG* and *yecS*).

A large proportion of the 52 candidate pathoadaptive genes have previously been associated with cystic fibrosis–associated infections. For example, ten of the genes (*gyrA*, *gyrB*, *mexA*, *mexB*, *mexR*, *mexS*, *mexZ*, *nalD*, *nfxB* and *oprD*) have been shown to be involved in resistance against a range of antibiotics[27,30–33], such as β-lactams, quinolones, chloramphenicol, macrolides, aminoglycosides and penicillins, and eight of the genes (*bifA*, *lasR*, *morA*, *phaF*, *rbdA*, *retS*, *wspA* and *wspE*) are involved in regulation of biofilm formation[34–40].

Yet, the implications for pathogenesis of other candidate pathoadaptive genes are less apparent. For example, 11 clone types accumulated nonsynonymous mutations (3 frameshifting indels, 1 in-frame indel, 10 missense mutations and 1 nonsense mutation) in either *aceE* or *aceF* encoding the E1 and E2 components, respectively, of the pyruvate dehydrogenase complex, which controls the flux of glycolytic carbon entering the tricarboxylic acid (TCA) cycle. Analyses of *P. aeruginosa* growth on a medium designed to represent the nutritional content of sputum from patients with cystic fibrosis have shown that *aceE* and *aceF* knockout strains produce large amounts of pyruvate[41], and, under anaerobic growth conditions (and in the absence of alternative electron acceptors), *P. aeruginosa* uses the conversion of pyruvate to acetate and lactate for long-term survival[42]. This observation suggests that the remodeling of central metabolism signaling has a role in the pathogenesis of *P. aeruginosa* within the host.

Finally, we also identified genes of unknown function and without previous implication in pathogenesis, for example, the PA3290 gene, which were among the 13 genes with the highest frequency of nonsynonymous mutations (**Fig. 5b**).

Convergence and constraints in genetic adaptation

Repeated nonsynonymous mutation of the same genes suggests that evolution toward some common beneficial phenotype is constrained to mutations in a single or a few genes. The *wsp* signal transduction pathway encoded by *wspABCDEFR* regulates biofilm formation through the modulation of cyclic diguanylate levels[40]. Although none of the 7 pathway genes were mutated in more than 7 clone types, we found that 14 of the clone types had acquired nonsynonymous mutations in at least one of the *wsp* genes. This finding indicates that there is strong selection for mutation of this regulatory network and that the adaptive phenotype is obtainable through multiple different evolutionary pathways.

In contrast, the mutational routes to some beneficial phenotypes might be more constrained. Accordingly, we found that the most frequently mutated gene, *mexZ*, had nonsynonymous mutations (24 frameshifting indels, 6 in-frame indels and 10 missense SNPs) in more than half (n = 20) of the 36 lineages (**Fig. 5b**), indicating that mutation of *mexZ* is part of a strongly selected but constrained evolutionary pathway. The MexZ protein is a negative regulator of the MexXY-OprM multidrug efflux pump[43,44], whose overexpression is associated with resistance to the aminoglycoside tobramycin[45], one of the first-line antibiotics used in the cystic fibrosis clinic, and *mexZ* has previously been reported to be frequently mutated during cystic fibrosis airway infections[2,46].

Contingency between mutations in signaling pathways

Constitutive or conditional overproduction of alginate leading to mucoid colony morphologies is another hallmark of cystic

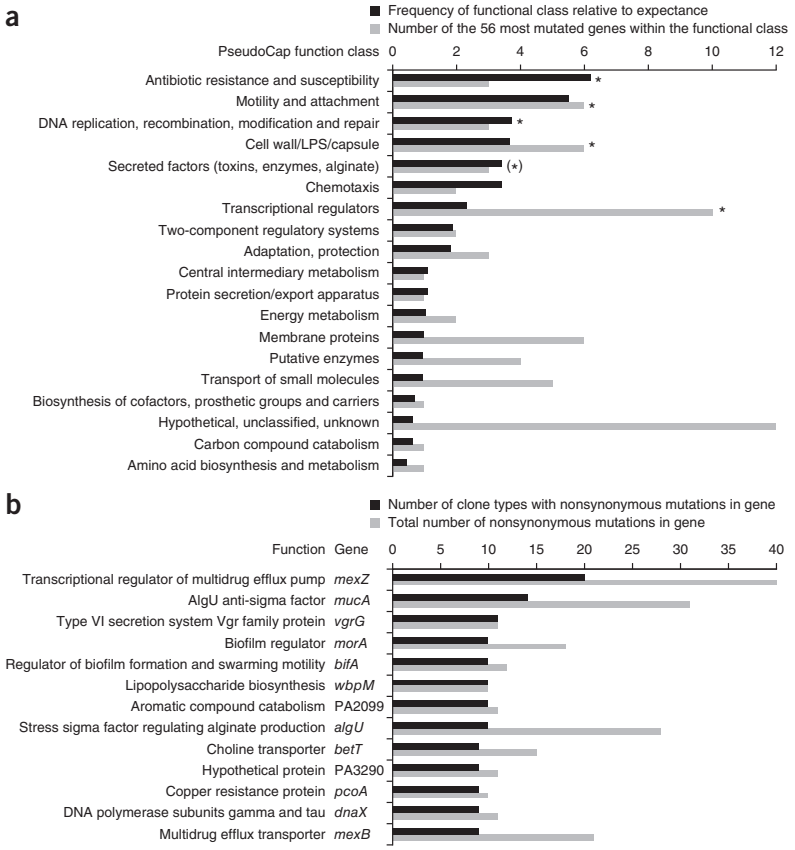


Figure 5 The most frequently mutated functional classes and genes. (**a**) Distribution on the PseudoCap functional classes of the genes ($n = 52$) subject to convergent evolution. Asterisks denote functional classes that are significantly over-represented among the 52 mutated genes ($P(X \geq x)$ ~binom(X; p) < 0.05, where $P(X \geq x)$ is the probability of observing $\geq x$ of the 52 genes belonging to a functional class present in the genome with a frequency p; **Supplementary Table 6**). The asterisk in parentheses indicates the P value for secreted factors (toxins, enzymes, alginate), 0.058. (**b**) Genes ($n = 13$) mutated in at least 9 different clone types. Horizontal bars show the number of mutated clonal types and the total number of nonsynonymous mutations (**Supplementary Table 5**).

fibrosis–associated infections[14]. Alginate production is positively regulated by the alternative sigma factor AlgU (also known as σ^{22} or σ^{E})[47,48], which is normally sequestered by binding to the anti-sigma factor MucA (**Fig. 6a**), and we found both of the corresponding genes to be among the most mutated candidate pathoadaptive genes, with 28 and 31 nonsynonymous mutations, respectively (**Fig. 5b**). There was a strong co-occurrence of mutations in *algU* and *mucA*, and 25 of the 28 *algU* mutations were found in isolates that also carried mutations in *mucA*. Intriguingly, the order in which the *algU* and *mucA* mutations occurred was not random, as the mutations in *algU* were consecutive to the mutations in *mucA* in 20 of the 25 *mucA* and *algU* double mutants (in the remaining 5 cases, phylogenetic inference could not be used to resolve the order of the mutations) (**Fig. 6b**). This finding constitutes an example of historical contingency, where a constitutive mucoid phenotype caused by loss-of-function mutation in *mucA* is fully or partly abolished by secondary mutation in *algU*. We suggest that this evolutionary trajectory is selected for in cystic fibrosis airways because constitutive mucoidity may have immediate benefits; however, regulation of this energy-draining phenotype may subsequently be remodeled by mutation of *algU*[23,48–50]. Because AlgU is also a positive regulator of the stress response[47,48], another possibility might be that a constitutive stress response, not mucoidity, is selected for by mutation of *mucA*. In this way, mucoidity might just be a pleiotropic effect that is compensated for by subsequent mutation of *algU*.

We searched our data set for other mutational dependencies, finding 11 cases in which the infecting lineages had acquired more than one nonsynonymous mutation in the *retS-gacS-gacA-rsmZ-rsmA* signaling pathway since the MRCA (**Supplementary Fig. 6**). The *retS-gacS-gacA-rsmZ-rsmA* signaling pathway reciprocally controls the expression of genes important for acute and chronic infection[39] (**Fig. 6c**), so we anticipated that mutations in this pathway might constitute another example of historical contingency.

RetS inhibits GacA/GacS-activated expression of the small regulatory RNA *rsmZ*, which binds to and sequesters RsmA. RsmA is responsible for the reciprocal regulation of genes important for acute and chronic infection (**Fig. 6c**). In all 11 infecting lineages, the upstream regulator *retS* was mutated before mutation of the downstream regulators *gacA*, *gacS* or *rsmA* (**Fig. 6d** and

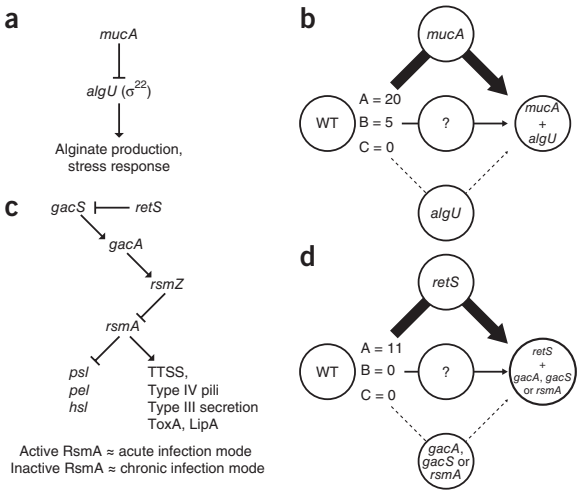


Figure 6 The order of mutations in mutants with two nonsynonymous mutations in the same regulatory pathway. (**a**–**d**) We identified 25 and 11 unique double mutants with mutations in 2 of the genes in the *mucA-algU* (**a**) and *retS-gacS-gacA-rsmZ-rsmA* (**c**) regulatory pathways, respectively. On the basis of the phylogeny of the mutants, we were able to infer the order of the mutations in the *mucA-algU* (**b**) and *retS-gacS-gacA-rsmZ-rsmA* (**d**) regulatory pathways, that is, which of the possible mutational routes (A, B or C) led to the double mutant. WT, wild type.

Supplementary Fig. 6). The mutational signatures of *retS*, *gacA* and *gacS* suggest that these genes are targeted by loss-of-function mutations (all three genes had frameshift mutations), whereas the two missense substitutions in RsmA were located in the RNA-binding site[51], thus preventing binding by *rsmZ*. Using the regulatory model devised by Goodman and colleagues[39] (**Fig. 6c**), we predicted that the *retS* mutations directed the bacteria toward a chronic infection state, whereas the secondary mutations in either *gacA*, *gacS* or *rsmA* reverted this effect. Furthermore, our findings suggest that, although the *retS-gacS-gacA-rsmZ-rsmA* signaling pathway seems to be important for infection, it might be overly simplifying to regard the regulatory pathway as a bimodal switch between either an acute or chronic infection state.

DISCUSSION

By sequencing and analyzing the genomes of 474 isolates of *P. aeruginosa*, we have gained insight into the evolution of an opportunistic pathogen after its environmental transition into the airways of patients with cystic fibrosis. To our knowledge, this is the largest and most diverse collection of genome-sequenced bacterial isolates from patients with cystic fibrosis. Although there are already a few other studies encompassing several hundreds of whole genome–sequenced isolates of bacterial pathogens[5–12], our study is unique in its focus on an opportunistic pathogen causing long-term infections. Initial bacterial isolates were sampled at the onset of infection in 34 young patients with cystic fibrosis, and sampling was continued, on average, for 4.8 years (~20,000 bacterial generations[52]), allowing us to obtain a comprehensive picture of the genetic adaptation of microorganisms to a new environment[4,53].

In addition, whereas many other studies are focusing on single clonal lineages or complexes of a pathogenic species, our collection compares the within-host evolution of 36 genotypically different strains of the same species. This is an important issue, as mutations or genes associated with pathogenicity in one strain of *P. aeruginosa* might not be predictive of pathogenicity in other strains[54]. We found evidence for convergent molecular evolution in 52 genes across the 36 lineages, with the majority having annotated functions related to phenotypes considered important for infection. Mutation of five of these genes (*lasR*, *mexA*, *mexS*, *nexZ* and *yecS*) was suggested to be involved in host adaptation in a study by Smith *et al.* that carried out the first genome comparison of longitudinally collected clinical isolates of *P. aeruginosa*[2]. Furthermore, another 7 of the 52 genes (*algU*, *gyrA*, *gyrB*, *mexB*, *oprD*, *pelA* and *rbdA*) are homologous to 65 genes previously reported to be important for cystic fibrosis–associated infections in our earlier study that followed the genomic evolution of the *P. aeruginosa* DK02 lineage as it disseminated through a cohort of adult patients with cystic fibrosis between the years 1972 and 2010 (ref. 1).

However, in this previous study, we were limited in our ability to identify mutations that are of immediate importance for *P. aeruginosa* upon transition from the environmental reservoir to cystic fibrosis airways, as such events might only have occurred once in the history of the DK02 lineage. Because the current study is conducted on young patients with cystic fibrosis who had been infected with *P. aeruginosa* for the first time, we anticipate that our investigation will facilitate a better understanding of the genetic adaptation that occurs right at the initial colonization of cystic fibrosis airways. In support of this hypothesis, we find that the 52 candidate pathoadaptive genes identified by us are more frequently mutated in the earlier stages of DK02 evolution (mutations accumulated before 1979) relative to the later stages (15 of 173 versus 109 of 4,962 nonsynonymous mutations, respectively; Fisher's exact test, $P = 1.4 \times 10^{-5}$).

The congruence between our suggested list of candidate pathoadaptive genes and genes found by others to be involved in host adaptation demonstrates the validity of our conclusion that different genetic backgrounds of an opportunistic pathogen show convergent adaptive evolution upon transition from the environment to human airways.

Nonetheless, it is also notable that we identified convergent molecular evolution in genes of unknown function or without previous implication in pathogenesis. Further investigations of the function of these genes are required to determine their potential as future therapeutic targets against infection. Furthermore, continued characterization of pathoadaptive mutations will help to link genotype to phenotype, and such knowledge will be valuable for clinicians with regard to the treatment and segregation of patients.

Cystic fibrosis airways constitute a complex environment with both spatial and temporal heterogeneity, allowing for the presence of multiple niches. This means that infecting populations may genetically diversify to establish sublineages, each genetically adapting under the selection pressure of its niche. Further studies are necessary to address whether some of the pathoadaptive mutations are linked to certain niches in the airways. Also, our results may in general be complemented by studies that take population heterogeneity into consideration when analyzing the genomic evolution of pathogens[55].

The role of historical contingency in evolution is challenging to assess in natural systems in which one cannot rewind the tape of evolution and replay it. Nonetheless, we demonstrated how the order of mutations in two different regulatory networks is highly constrained, and, as a result, mutations in downstream transcriptional regulators were contingent upon mutations in upstream regulators. Intriguingly, initial mutation of these networks directs the bacteria toward a chronic infection state (for example, with increased production of biofilm-promoting polysaccharides), but a secondary mutation followed to interfere with the effect of the first mutation. The secondary mutations might rebalance the regulatory network to the wild-type status quo, or, alternatively, the historical contingency patterns exemplify evolutionary adaptation facilitated by remodeling of regulatory networks. In the latter case, the secondary mutation modulates the detrimental pleiotropic effects of the primary mutation of a global regulator[23]. We speculate that these evolutionary trajectories are the result of a complex fitness landscape in which multiple strictly ordered mutations are necessary to reach fitness peaks. Furthermore, we anticipate that knowledge of constraints on the order of mutations might help in the prediction of bacterial evolution in patients with cystic fibrosis and the design of future treatment strategies. For example, certain bacterial evolutionary trajectories may entail a fitness cost that can be targeted by treatment, or treatment may even be used to direct bacteria toward evolutionary trajectories that are associated with lower patient morbidity.

Finally, we found hypermutator strains in 5 of the 34 patients in our study (15%). This fraction is lower than the range of 36–54% that has been reported from four other systematic studies of hypermutator prevalence[1,20,56,57]. We ascribe this difference to the relatively young age of the patients in this study, as Ciufo *et al.* found no mutators until 5 years after the onset of chronic lung infection among the 79 patients with cystic fibrosis tested[56].

In conclusion, we have shown how a clinical collection of bacteria sampled from chronically infected patients constitutes a valuable basis for understanding the evolutionary convergence and contingency of pathogens *in vivo*. Our results facilitate comparative studies

as sequencing data sets become increasingly available and may help in the design of future intervention strategies for the clinical setting.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequence reads from all *P. aeruginosa* isolates have been deposited in the Sequence Read Archive (SRA) under accession ERP004853. See **Supplementary Table 7** for the accession codes for individual isolates.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank U.R. Johansen, P. Poss, H. Nordbjerg, N. Kirkby, K. Bloksted and B.H. Erichsen for excellent technical assistance and T. Pressler and M. Skov for information about the patients' visits to the hospital. This work was supported by the Lundbeck Foundation, and H.K.J. was supported by a clinical research stipend from the Novo Nordisk Foundation.

AUTHOR CONTRIBUTIONS

S.M. and H.K.J. jointly supervised the study. R.L.M., S.M. and H.K.J. conceived and designed the experiments. H.K.J. collected clinical samples and provided clinical information. L.M.S. prepared the genomic libraries for whole-genome sequencing. R.L.M. designed the bioinformatics workflows for the analysis. R.L.M. and L.M.S. conducted whole-genome sequence analysis. R.L.M., L.M.S., S.M. and H.K.J. analyzed and interpreted the data. R.L.M. wrote the manuscript. L.M.S., S.M. and H.K.J. helped write the manuscript and provided revisions.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Marvig, R.L., Johansen, H.K., Molin, S. & Jelsbak, L. Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet.* **9**, e1003741 (2013).
2. Smith, E.E. *et al.* Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl. Acad. Sci. USA* **103**, 8487–8492 (2006).
3. Cramer, N. *et al.* Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs. *Environ. Microbiol.* **13**, 1690–1704 (2011).
4. Yang, L. *et al.* Evolutionary dynamics of bacteria in a human host environment. *Proc. Natl. Acad. Sci. USA* **108**, 7481–7486 (2011).
5. Holt, K.E. *et al.* Tracking the establishment of local endemic populations of an emergent enteric pathogen. *Proc. Natl. Acad. Sci. USA* **110**, 17522–17527 (2013).
6. Eyre, D.W. *et al.* Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N. Engl. J. Med.* **369**, 1195–1205 (2013).
7. Chewapreecha, C. *et al.* Dense genomic sampling identifies highways of pneumococcal recombination. *Nat. Genet.* **46**, 305–309 (2014).
8. Comas, I. *et al.* Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* **45**, 1176–1182 (2013).
9. Casali, N. *et al.* Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat. Genet.* **46**, 279–286 (2014).
10. Mather, A.E. *et al.* Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science* **341**, 1514–1517 (2013).
11. Grad, Y.H. *et al.* Genomic epidemiology of *Neisseria gonorrhoeae* with reduced susceptibility to cefixime in the USA: a retrospective observational study. *Lancet Infect. Dis.* **14**, 220–226 (2014).
12. Croucher, N.J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434 (2011).
13. Sokurenko, E.V., Hasty, D.L. & Dykhuizen, D.E. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol.* **7**, 191–195 (1999).
14. Folkesson, A. *et al.* Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: an evolutionary perspective. *Nat. Rev. Microbiol.* **10**, 841–851 (2012).
15. Johansen, H.K., Moskowitz, S.M., Ciofu, O., Pressler, T. & Hoiby, N. Spread of colistin resistant non-mucoid *Pseudomonas aeruginosa* among chronically infected Danish cystic fibrosis patients. *J. Cyst. Fibros.* **7**, 391–397 (2008).
16. Wiehlmann, L. *et al.* Population structure of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA* **104**, 8101–8106 (2007).
17. Jelsbak, L. *et al.* Molecular epidemiology and dynamics of *Pseudomonas aeruginosa* populations in lungs of cystic fibrosis patients. *Infect. Immun.* **75**, 2214–2224 (2007).
18. Zimakoff, J., Hoiby, N., Rosendal, K. & Guilbert, J.P. Epidemiology of *Pseudomonas aeruginosa* infection and the role of contamination of the environment in a cystic fibrosis clinic. *J. Hosp. Infect.* **4**, 31–40 (1983).
19. Worby, C.J., Lipsitch, M. & Hanage, W.P. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLoS Comput. Biol.* **10**, e1003549 (2014).
20. Oliver, A., Canton, R., Campo, P., Baquero, F. & Blazquez, J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288**, 1251–1254 (2000).
21. Winsor, G.L. *et al.* *Pseudomonas* Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes. *Nucleic Acids Res.* **39**, D596–D600 (2011).
22. Luzar, M.A. & Montie, T.C. Avirulence and altered physiological properties of cystic fibrosis strains of *Pseudomonas aeruginosa*. *Infect. Immun.* **50**, 572–576 (1985).
23. Damkiær, S., Yang, L., Molin, S. & Jelsbak, L. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proc. Natl. Acad. Sci. USA* **110**, 7766–7771 (2013).
24. Pai, H. *et al.* Carbapenem resistance mechanisms in *Pseudomonas aeruginosa* clinical isolates. *Antimicrob. Agents Chemother.* **45**, 480–484 (2001).
25. Ballestero, S. *et al.* Carbapenem resistance in *Pseudomonas aeruginosa* from cystic fibrosis patients. *J. Antimicrob. Chemother.* **38**, 39–45 (1996).
26. Schurek, K.N. *et al.* Involvement of *pmrAB* and *phoPQ* in polymyxin B adaptation and inducible resistance in non-cystic fibrosis clinical isolates of *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **53**, 4345–4351 (2009).
27. Cabot, G. *et al.* Genetic markers of widespread extensively drug-resistant *Pseudomonas aeruginosa* high-risk clones. *Antimicrob. Agents Chemother.* **56**, 6349–6357 (2012).
28. Juan, C. *et al.* Molecular mechanisms of β-lactam resistance mediated by AmpC hyperproduction in *Pseudomonas aeruginosa* clinical strains. *Antimicrob. Agents Chemother.* **49**, 4733–4738 (2005).
29. Mahenthiralingam, E., Campbell, M.E. & Speert, D.P. Nonmotility and phagocytic resistance of *Pseudomonas aeruginosa* isolates from chronically colonized patients with cystic fibrosis. *Infect. Immun.* **62**, 596–605 (1994).
30. Sobel, M.L., Hocquet, D., Cao, L., Plesiat, P. & Poole, K. Mutations in PA3574 (*nalD*) lead to increased MexAB-OprM expression and multidrug resistance in laboratory and clinical isolates of *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **49**, 1782–1786 (2005).
31. Hancock, R.E. Resistance mechanisms in *Pseudomonas aeruginosa* and other nonfermentative Gram-negative bacteria. *Clin. Infect. Dis.* **27** (suppl. 1), S93–S99 (1998).
32. Strateva, T. & Yordanov, D. *Pseudomonas aeruginosa*—a phenomenon of bacterial resistance. *J. Med. Microbiol.* **58**, 1133–1148 (2009).
33. Pasca, M.R. *et al.* Evaluation of fluoroquinolone resistance mechanisms in *Pseudomonas aeruginosa* multidrug resistance clinical isolates. *Microb. Drug Resist.* **18**, 23–32 (2012).
34. Huse, H.K. *et al.* *Pseudomonas aeruginosa* enhances production of a non-alginate exopolysaccharide during long-term colonization of the cystic fibrosis lung. *PLoS ONE* **8**, e82621 (2013).
35. Kuchma, S.L. *et al.* BifA, a cyclic-Di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. *J. Bacteriol.* **189**, 8165–8178 (2007).
36. Davies, D.G. *et al.* The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* **280**, 295–298 (1998).
37. Choy, W.K., Zhou, L., Syn, C.K., Zhang, L.H. & Swarup, S. MorA defines a new class of regulators affecting flagellar development and biofilm formation in diverse *Pseudomonas* species. *J. Bacteriol.* **186**, 7221–7228 (2004).
38. An, S., Wu, J. & Zhang, L.H. Modulation of *Pseudomonas aeruginosa* biofilm dispersal by a cyclic-Di-GMP phosphodiesterase with a putative hypoxia-sensing domain. *Appl. Environ. Microbiol.* **76**, 8160–8173 (2010).
39. Goodman, A.L. *et al.* A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in *Pseudomonas aeruginosa*. *Dev. Cell* **7**, 745–754 (2004).
40. Hickman, J.W., Tifrea, D.F. & Harwood, C.S. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc. Natl. Acad. Sci. USA* **102**, 14422–14427 (2005).
41. Behrends, V. *et al.* Metabolite profiling to characterize disease-related bacteria: gluconate excretion by *Pseudomonas aeruginosa* mutants and clinical isolates from cystic fibrosis patients. *J. Biol. Chem.* **288**, 15098–15109 (2013).
42. Eschbach, M. *et al.* Long-term anaerobic survival of the opportunistic pathogen *Pseudomonas aeruginosa* via pyruvate fermentation. *J. Bacteriol.* **186**, 4596–4604 (2004).
43. Aires, J.R., Kohler, T., Nikaido, H. & Plesiat, P. Involvement of an active efflux system in the natural resistance of *Pseudomonas aeruginosa* to aminoglycosides. *Antimicrob. Agents Chemother.* **43**, 2624–2628 (1999).
44. Westbrock-Wadman, S. *et al.* Characterization of a *Pseudomonas aeruginosa* efflux pump contributing to aminoglycoside impermeability. *Antimicrob. Agents Chemother.* **43**, 2975–2983 (1999).
45. Sobel, M.L., McKay, G.A. & Poole, K. Contribution of the MexXY multidrug transporter to aminoglycoside resistance in *Pseudomonas aeruginosa* clinical isolates. *Antimicrob. Agents Chemother.* **47**, 3202–3207 (2003).

46. Islam, S., Jalal, S. & Wretlind, B. Expression of the MexXY efflux pump in amikacin-resistant isolates of *Pseudomonas aeruginosa*. *Clin. Microbiol. Infect.* **10**, 877–883 (2004).
47. Yu, H., Schurr, M.J. & Deretic, V. Functional equivalence of *Escherichia coli* σ^E and *Pseudomonas aeruginosa* AlgU: *E. coli rpoE* restores mucoidy and reduces sensitivity to reactive oxygen intermediates in *algU* mutants of *P. aeruginosa*. *J. Bacteriol.* **177**, 3259–3268 (1995).
48. DeVries, C.A. & Ohman, D.E. Mucoid-to-nonmucoid conversion in alginate-producing *Pseudomonas aeruginosa* often results from spontaneous mutations in *algT*, encoding a putative alternate sigma factor, and shows evidence for autoregulation. *J. Bacteriol.* **176**, 6677–6687 (1994).
49. Schurr, M.J., Martin, D.W., Mudd, M.H. & Deretic, V. Gene cluster controlling conversion to alginate-overproducing phenotype in *Pseudomonas aeruginosa*: functional analysis in a heterologous host and role in the instability of mucoidy. *J. Bacteriol.* **176**, 3375–3382 (1994).
50. Ciofu, O. *et al.* Investigation of the *algT* operon sequence in mucoid and non-mucoid *Pseudomonas aeruginosa* isolates from 115 Scandinavian patients with cystic fibrosis and in 88 *in vitro* non-mucoid revertants. *Microbiology* **154**, 103–113 (2008).
51. Heeb, S. *et al.* Functional analysis of the post-transcriptional regulator RsmA reveals a novel RNA-binding site. *J. Mol. Biol.* **355**, 1026–1036 (2006).
52. Yang, L. *et al. In situ* growth rates and biofilm development of *Pseudomonas aeruginosa* populations in chronic lung infections. *J. Bacteriol.* **190**, 2767–2776 (2008).
53. Barrick, J.E. *et al.* Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247 (2009).
54. Lee, D.G. *et al.* Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol.* **7**, R90 (2006).
55. Lieberman, T.D. *et al.* Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat. Genet.* **46**, 82–87 (2014).
56. Ciofu, O., Riis, B., Pressler, T., Poulsen, H.E. & Hoiby, N. Occurrence of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. *Antimicrob. Agents Chemother.* **49**, 2276–2282 (2005).
57. Waine, D.J., Honeybourne, D., Smith, E.G., Whitehouse, J.L. & Dowson, C.G. Association between hypermutator phenotype, clinical variables, mucoid phenotype, and antimicrobial resistance in *Pseudomonas aeruginosa*. *J. Clin. Microbiol.* **46**, 3491–3493 (2008).

ONLINE METHODS

Bacterial isolates. This study included 474 clinical isolates of *P. aeruginosa* that were sampled from 34 patients with cystic fibrosis attending the Copenhagen Cystic Fibrosis Center at the University Hospital, Rigshospitalet, Denmark. Isolation and identification of *P. aeruginosa* from sputum was carried out as previously described[58]. Use of the samples was approved by the local ethics committee of the Capital Region of Denmark (Region Hovedstaden; registration numbers H-A-141 and H-1-2013-032), and all patients gave informed consent.

Genome sequencing. Genomic DNA was prepared from *P. aeruginosa* isolates on a QIAcube system using a DNeasy Blood and Tissue kit (Qiagen) and sequenced on an Illumina HiSeq 2000 platform, generating 100-bp paired-end reads and using a multiplexed protocol to obtain an average of 7,139,922 reads (range of 3,111,062–13,085,190) for each of the genomic libraries. On average, we sequenced the isolates with an estimated genomic coverage of 107-fold (range of 55- to 195-fold). See **Supplementary Table 7** for information about genomic coverage depths for individual isolates.

Definition of clone types. Sequence reads from each isolate were *de novo* assembled using Velvet[59] (version 1.2.03) with a *k*-mer length of 33 and the options set as follows: '-scaffolding no -ins_length 500 -cov_cutoff 3 -min_contig_lgth 300'. *De novo*–assembled genomes were aligned against each other using MUMmer3 (ref. 60) (version 3.23), and SNPs bounded by 20 exact base-pair matches on both sides were extracted from the alignment. Isolates with genomes that differed by >10,000 SNPs were considered to belong to different clone types. When considering only SNPs bounded by 50 exact base-pair matches on both sides, genomes belonging to different clone types differed by >6,000 SNPs.

Mutation detection and analysis. Mutations were identified as described in Marvig *et al.*[1] with the following modifications. We aligned reads to the *P. aeruginosa* PAO1 reference genome (GenBank accession NC_002516.2; genome size of 6.4 Mb) with Bowtie 2 (ref. 61), used the Genome Analysis Toolkit (GATK) for realignment around indels[62] and produced pileups of the read alignments with SAMtools release 0.1.7 (ref. 63).

SNPs were extracted from the read pileup if they met the following criteria: (i) a quality score (Phred-scaled probability of the sample reads being homozygous reference) of at least 50 (i.e., $P \leq 1 \times 10^{-5}$), (ii) a root-mean-square (RMS) mapping quality of at least 25, (iii) a minimum of three reads covering the position and (iv) only unambiguous SNP calls. Microindels were extracted from the read pileup if they met the following criteria: (i) a quality score of at least 500, (iii) an RMS mapping quality of at least 25 and (iii) support from at least one-fifth of the covering reads.

To avoid false positives caused by strain-specific (native) differences relative to the published genome sequence of *P. aeruginosa* strain PAO1, we excluded polymorphisms specific to the genetic background. This means that polymorphisms shared by all members of a clone type were excluded from the analysis; that is, only mutations that had accumulated since the MRCA of the clonal isolates were included. Only genomic positions covered by at least three reads in all members of a clone type were included in the analysis.

Maximum-parsimonious phylogenetic analyses were carried out with PAUP* version 4.0b10 (ref. 64) using alleles of reference strain PAO1 as a root. Consistency indexes (CI = m/s) were calculated as the number of mutations (m) divided by the minimum number of mutational events required to explain the phylogenies (s). The CI equals 1 when there is no homoplasy.

Identification of candidate pathoadaptive genes. To identify significant patterns of convergent evolution of nonsynonymous mutations across multiple clone types, we evaluated the observed distribution of mutations considering a scenario in which mutations were randomly introduced by genetic drift. For each clone type, we randomly selected m positions in the coding part of the *P. aeruginosa* reference genome, where m was the number of nonsynonymous mutations in the respective clone type. We then recorded the genes that were randomly targeted by mutation and repeated the procedure 1,000 times. For each of the 1,000 rounds of mutation, we counted the number of genes that were mutated in $\geq x$ clone types. As larger genes are more likely to be randomly mutated, we grouped the results on the basis of the lengths of the genes. Finally, we compared the observed counts of nonsynonymous mutations with the distribution of nonsynonymous mutations expected by genetic drift, and, whenever the observed number of mutations for a gene was significantly higher ($P \leq 0.0011$) and at least tenfold higher relative to expectance, the gene was added to the list of candidate pathoadaptive genes (**Supplementary Table 4**). Accordingly, genes in the size range of 1–1,000 nt were listed when mutated in ≥5 clone types, genes in the size range of 1,001–3,000 nt were listed when mutated in ≥7 clone types and genes in the size range of 3,001–5,000 nt were listed when mutated in ≥9 clone types.

The counts of the clone types mutated in the candidate pathoadaptive genes were not affected by removing densely clustered mutations (mutations within 50 bp of each other) that were linked in the phylogenetic reconstruction, for example, because of recombination events.

Antibiotic treatment of patients. Early and systematic antibiotic chemotherapy is used to treat patients for *P. aeruginosa*, and 33 of the 34 patients in the study were treated with fluoroquinolone (ciprofloxacin), aminoglycoside (tobramycin), macrolide (azithromycin), β-lactam (piperacillin, meropenem, ceftazidime and/or aztreonam) and antimicrobial peptide (colistin) antibiotics. The only exception was patient P51M5 who was not treated with a macrolide antibiotic. Patients were treated according to the following guidelines.

Treatment was initiated at first detection of *P. aeruginosa* through the administration of oral ciprofloxacin in combination with nebulized colistin for 3 months (or for 3 weeks, beginning in 2008). If *P. aeruginosa* still appeared in sputum samples after 3 months (weeks), intravenous treatment with piperacillin or tazobactam in combination with tobramycin was given (tobramycin might be inhaled in some cases). If more than 3 months (weeks) elapsed before *P. aeruginosa* reappeared, treatment with ciprofloxacin and colistin was restarted. When piperacillin or tazobactam resistance was observed, meropenem, ceftazidime or aztreonam might be used in combination with tobramycin or colistin. **Supplementary Table 8** lists the antibiotics used to treat each of the patients for *P. aeruginosa*.

58. Hoiby, N. & Frederiksen, B. in *Cystic Fibrosis* (eds. Hodson, M. & Geddes, D.) 83–107 (Arnold, London, 2000).
59. Zerbino, D.R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
60. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
61. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
62. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
63. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. Swofford, D.L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4* (Sinauer Associates, Sunderland, MA, 2003).

4.3 **study 3:**

Within-host microevolution of *Pseudomonas aeruginosa* in Italian cystic fibrosis patients.

Marvig R. L.*, Dolce D.*, **Sommer L. M.***, Petersen B., Ciofu O., Campana S., Molin S., Taccetti G., and Johansen H. K. (2015), *Submitted for publication*

Marvig et al. BMC Microbiology (2015) 15:218
DOI 10.1186/s12866-015-0563-9

BMC Microbiology

RESEARCH ARTICLE Open Access

Within-host microevolution of *Pseudomonas aeruginosa* in Italian cystic fibrosis patients

Rasmus Lykke Marvig[1,2*†], Daniela Dolce[3†], Lea M. Sommer[4,5†], Bent Petersen[6], Oana Ciofu[1], Silvia Campana[3], Søren Molin[4,5], Giovanni Taccetti[3] and Helle Krogh Johansen[1,5]

Abstract

Background: Chronic infection with *Pseudomonas aeruginosa* is a major cause of morbidity and mortality in cystic fibrosis (CF) patients, and a more complete understanding of *P. aeruginosa* within-host genomic evolution, transmission, and population genomics may provide a basis for improving intervention strategies. Here, we report the first genomic analysis of *P. aeruginosa* isolates sampled from Italian CF patients.

Results: By genome sequencing of 26 isolates sampled over 19 years from four patients, we elucidated the within-host evolution of clonal lineages in each individual patient. Many of the identified mutations were located in pathoadaptive genes previously associated with host adaptation, and we correlated mutations with changes in CF-relevant phenotypes such as antibiotic resistance. In addition, the genomic analysis revealed that three patients shared the same clone. Furthermore, we compared the genomes of the Italian CF isolates to a panel of genome sequenced strains of *P. aeruginosa* from other countries. Isolates from two of the Italian lineages belonged to clonal complexes of *P. aeruginosa* that have previously been identified in Danish CF patients, and our genomic comparison showed that clonal isolates from the same country may be more distantly related than clonal isolates from different countries.

Conclusions: This is the first whole-genome analysis of *P. aeruginosa* isolated from Italian CF patients, and together with both phenotypic and clinical information this dataset facilitates a more detailed understanding of *P. aeruginosa* within-host genomic evolution, transmission, and population genomics. We conclude that the evolution of the Italian lineages resembles what has been found in other countries.

Keywords: Bacterial pathogens, Genetic adaptation, Evolution, Cystic fibrosis

Background

Advances in high-throughput DNA sequencing techniques have made it possible to follow the within-host genomic evolution of bacterial pathogens by comparing genomes of longitudinally collected bacterial isolates sampled from human hosts [1]. The genomic information may be used to understand pathogen population diversity, host adaptation, and routes and sources of transmission.

Pseudomonas aeruginosa infections in cystic fibrosis (CF) patients represents an infectious disease scenario in which within-host genomic evolution of clonal lineages of infecting bacteria can be followed by making genomic comparisons of clonal isolates sampled over time [2]. A number of clinical collections of freeze-stored *P. aeruginosa* isolates from CF patients have been genome sequenced to investigate pathogen microevolution. This includes investigation of within-host evolution of *P. aeruginosa* lineages isolated from CF patients from Argentina, Canada, Denmark, Germany, United Kingdom, and United States, respectively [3–10]. Nonetheless, to obtain a more complete basis for understanding *P. aeruginosa* infections in CF patients and to facilitate comparative studies to assess the generality of findings, it is necessary to investigate more clinical collections of *P. aeruginosa*. In order to serve as references, such investigations should be as comprehensive as possible, for example by including both genomic, phenotypic, and clinical information.

* Correspondence: rmarvig@gmail.com
†Equal contributors
[1]Department of Immunology and Microbiology, Costerton Biofilm Center, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[2]Center for Genomic Medicine, Rigshospitalet, Copenhagen, Denmark
Full list of author information is available at the end of the article

Here, we report the first genomic analysis of *P. aeruginosa* isolates sampled from Italian CF patients, aiming to compare whether within-host evolution of *P. aeruginosa* in Italian CF patients resembles what has been found in other countries.

In total, we sequenced 26 isolates sampled over a period of 19 years from four different patients. Genomic comparisons showed that the 26 isolates belonged to six different clone types (clonal complexes), which we named IT01-IT06. Each clone type was found to infect a single patient, except for clone type IT05 that was shared among three patients.

We conducted both an inter-clonal genomic analysis to determine the genetic relationship between different strains, and an intra-clonal genomic analysis to reveal within-host microevolution of the individual clonal lineages. We correlated the genetic changes with changes in relevant phenotypes and patient treatment. Finally, we compared the findings from this study of evolution of *P. aeruginosa* in Italian CF patients to the findings from similar studies encompassing lineages and patient cohorts from other countries. Altogether, continued genome sequencing of bacterial pathogens will improve our understanding of within-host genomic evolution, transmission, and population genomics.

Results

Collection of *P. aeruginosa* isolates from Italian CF patients

In order to investigate the within-host genomic evolution of *P. aeruginosa* lineages infecting CF patients, we performed a retrospective study of a collection of *P. aeruginosa* isolates from four patients attending the Cystic Fibrosis Center at Anna Meyer's Children University Hospital (Florence, Italy). The four patients named E, F, H and L were all born in the beginning of the 1990s (1991–1993), and they had their first colonization by *P. aeruginosa* within the first three years of life. The first and subsequent lung isolates of *P. aeruginosa* from the four patients were all stored to yield a collection of 35 isolates sampled over a time period of 19 years (1993 to 2012) (Fig. 1).

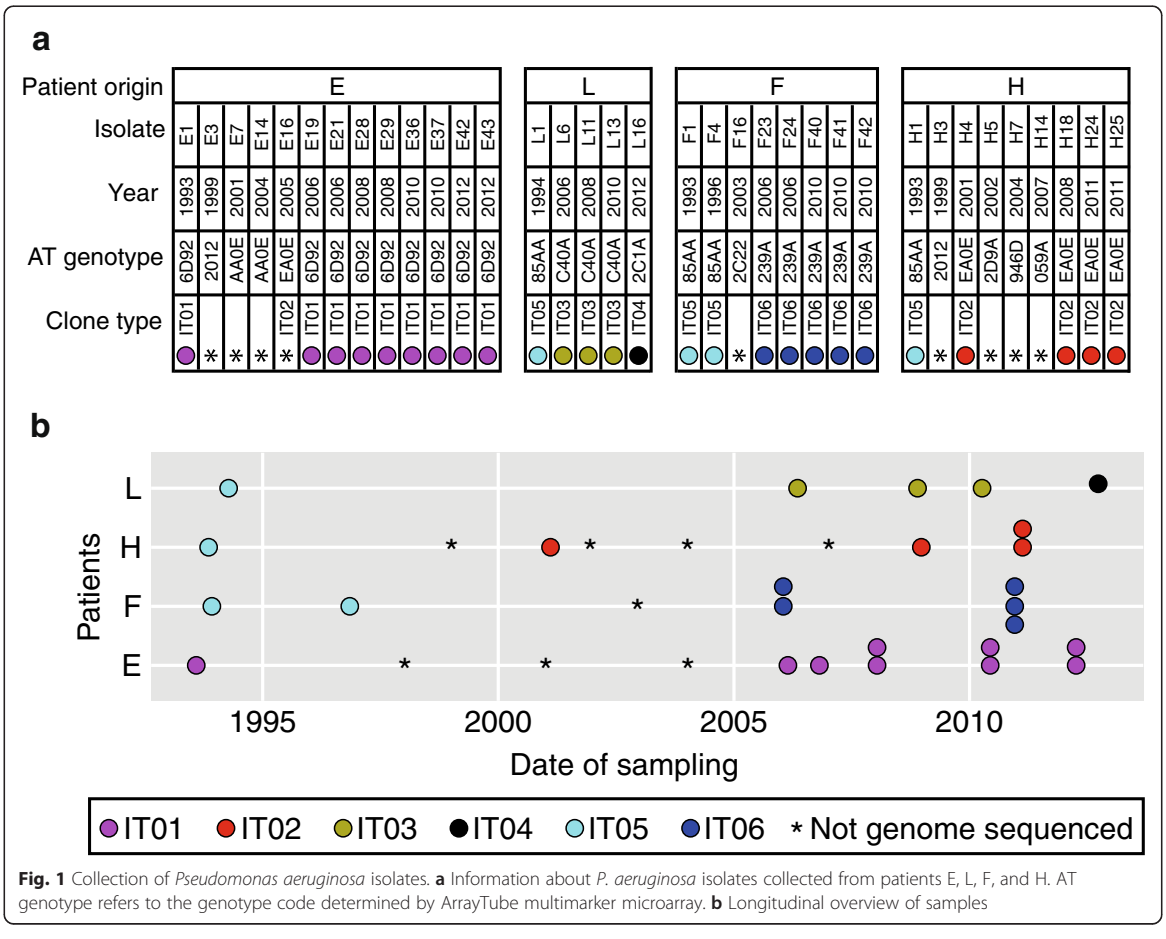


Fig. 1 Collection of *Pseudomonas aeruginosa* isolates. **a** Information about *P. aeruginosa* isolates collected from patients E, L, F, and H. AT genotype refers to the genotype code determined by ArrayTube multimarker microarray. **b** Longitudinal overview of samples

The ability of *P. aeruginosa* to cause infection in CF patients is not pertained to a single or a few strains (i.e. clonal complexes or clone types) of *P. aeruginosa*, and CF children are anticipated to acquire their first *P. aeruginosa* infection from environmental strains that are naïve to the human airways. Accordingly, many different and unique strains are observed to infect CF patients [9, 11–13]. While the early *P. aeruginosa* infections may possibly be eradicated, it is commonly observed that the same clone type persist in the airways of CF patients, and a single clone type is typically observed to be dominant [9]. In order to identify such persistent lineages, we typed the 35 strains using the ArrayTube multimarker microarray targeting 13 single nucleotide polymorphisms (SNPs) in the core genome and additional genetic markers in the accessory genome [13]. ArrayTube genotyping revealed that all four patients harbored reoccurring genotypes (Fig. 1). The reoccurrence of identical bacterial genotypes in the same patient likely reflect that strains persist in the airways of the patients [14]. Alternatively, the reoccurrence of identical genotypes may be due to independent re-colonization from the same environmental source.

Genome sequencing of isolates to determine genetic relationships

To further elucidate the possibility that *P. aeruginosa* lineages persisted within the airways of the patients, we genome sequenced 26 of the isolates to investigate their genomic relationships (Fig. 1). Genome sequencing confirmed that isolates of the same genotype were closely related (i.e. isolates are of the same clonal lineage), and we name these clonal lineages IT01, IT02, IT03, IT05, and IT06. Only one isolate of ArrayTube genotype 2C1A was sequenced, and we name this strain IT04.

Isolates of the same lineage differed at most by 84 SNPs (isolates E1 and E36 sampled 17 years apart), and on average 41 SNPs separated clonal isolates. Most diversity was present among isolates of the IT01 lineage (average genetic distance between IT01 isolates: 57 SNPs), whereas almost no SNP differences were observed between isolates belonging to lineage IT05 (isolates were separated by at most 2 SNPs).

Transmission of the IT05 lineage between patients

Lineages IT01, IT02, IT03, and IT06 were found to infect only a single patient (patients E, H, L, and F, respectively), suggesting that the patients had picked up the respective lineages from the environment. Contrary, lineage IT05 was found in both patients F, H, and L, suggesting that the IT05 lineage had spread among the patients by either direct patient-to-patient transmission or indirect transmission via a common environmental reservoir. Interestingly, the IT05 lineage is the first clone type of *P. aeruginosa* to be isolated from all three patients in years 1993, 1993, and 1994, respectively. However, in all cases other clone types of *P. aeruginosa* replace the IT05 clone type. We speculate that the IT05 lineage reside in a common environmental source to which the patients are frequently exposed, enabling the IT05 lineage to successfully colonize multiple patients, but that the IT05 lineage in the long run is replaced by other clone types that are inherently better to thrive in the patients' airways.

Phylogeny and mutational signatures

Next, we compared clonally related genomes to reconstruct the evolutionary history of each of the clonal lineages (Fig. 2). The mutations in each of the lineages accumulated in a highly parsimonious fashion (average parsimonious consistency of 0.99; Table 1), reflecting a unidirectional and clonal evolution since the most recent common ancestor. Thus, using a maximum-parsimonious phylogenetic model, we were able to make accurate inferences about the succession of mutations and the relationship among *P. aeruginosa* clones.

General conclusions about to what extent natural selection has been the driving force in the fixation of genetic variants, can be inferred from measuring the relative rates of nonsynonymous (dN) and synonymous (dS) genetic changes [15]. A dN/dS ratio greater than one implies that there has been an overall positive selection for mutations; whereas a ratio less than one implies that there has been a selection for removal of mutations, i.e. negative selection. However, note that the dN/dS ratio should be interpreted as an average signal of selection not giving any information on if both positive and negative selection has acted at different sites and/or times during the evolution. We found a significant signature of negative selection in lineages IT01 (dN/dS = 0.7) and IT03 (dN/dS = 0.3), whereas the dN/dS ratios in the other three lineages were neither significantly positive nor negative (Table 1). Our observed range of dN/dS from 0.3 to 1.2 is in accordance with previous findings of within-host microevolution to be affected by both negative [5, 8] and positive selection [10].

Furthermore, we performed Bayesian analyses of mutation rates, and we estimated the yearly rate of SNP mutations in IT01 lineage to be 2.7 SNPs/year (95 % highest posterior density (HPD; see Methods) 1.0–4.4 SNPs/year), which is equivalent to 4.5×10^{-7} SNPs/year per site. This means that the mutation rate of the IT01 lineage is almost identical to the within-host mutation rate (2.6 SNPs/year) estimated for the DK02 lineage evolving in chronically infected Danish CF patients [8]. Also, the rate is within the range of mutation rates estimated in a number of other studies: Snyder et al. reported a mutation rate of 2.4-3.0

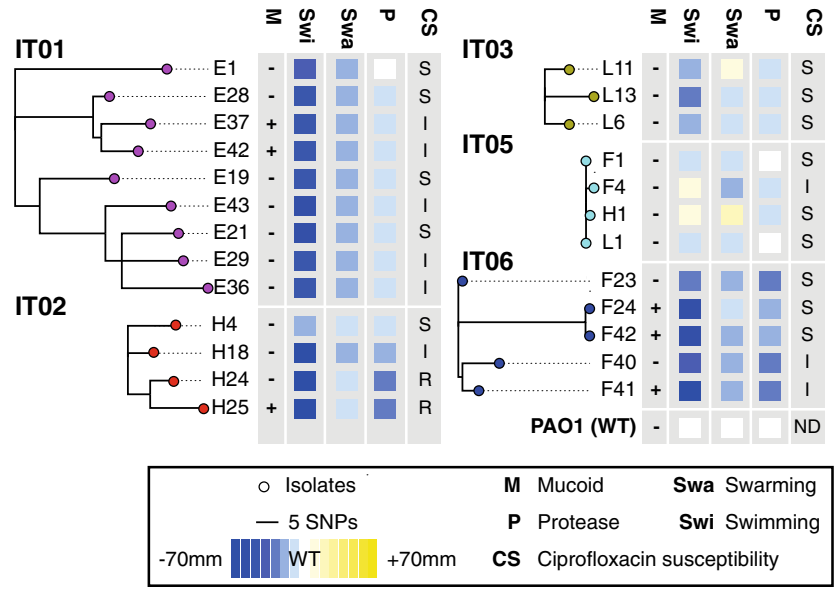


Fig. 2 Phylogeny and selected phenotypes of *Pseudomonas aeruginosa* lineages IT01, IT02, IT03, IT05, and IT06. Maximum-parsimonious phylogenetic trees were constructed by PAUP* (Methods). Size of swimming, swarming, and protease clearing zones relative to reference strain PAO1 are indicated by colors. Presence of mucoid phenotype is denoted by '+'. Ciprofloxacin susceptibility is indicated by 'S' (>21 mm clearing zone), 'I' (16–20 mm clearing zone), or 'R' (no clearing zone)

SNPs/year during a hospital outbreak [16]; Markussen et al. found the DK01 lineage to accumulate 1.3 SNPs/year in a chronically infected Danish CF patient; and Cramer et al. estimated the mutation rate of lineage PA14 to be ~1 SNP/year over the course of infection of a German CF patient [3]. Mutation rates of the other lineages (IT02, IT03, IT05, IT06) could not be estimated since the number and temporal distribution of isolates were insufficient to obtain proper estimates of the mutation rates (effective sample sizes (ESS) of modeled parameters were below 10).

Positive selection for mutations in pathoadaptive genes

Genetic adaptation is hypothesized to play a major role in the successful establishment of chronic *P. aeruginosa* infections of CF patients [3, 8–10, 17–22], and Marvig et al. recently performed a genome-wide mutational analysis of 36 different *P. aeruginosa* lineages to identify 52 candidate 'pathoadaptive genes' targeted by mutations to optimize pathogen fitness [9]. We found that 34 of 365 identified intragenic mutations were located within the 52 candidate pathoadaptive genes, corresponding to an 11-fold enrichment of mutations in

Table 1 Mutations accumulated in *Pseudomonas aeruginosa* lineages during infection

Lineage	IT01	IT02	IT03	IT05	IT06
SNPs	193	38	21	3	48
Missense	104	17	7	1	25
Nonsense	7	1	0	1	2
Silent	53	5	7	1	11
Intergenic	29	15	7	0	10
Intragenic indels	46	23	7	4	43
Intergenic indels	18	6	3	2	13
dN/dS	0.70	1.20	0.33	0.67	0.82
Probability of dN/dS = 1	3.61E-05	8.19E-01	3.95E-04	4.23E-01	1.70E-01
Maximum-parsimonious SNP events	193	39	21	3	48
Parsimony consistency	1.00	0.97	1.00	1.00	1.00

these particular 52 genes relative to all other genes ($P(X \geq 34) \sim pois(\lambda = 3.34) = 7.8 \times 10^{-22}$, where λ is the expected number of mutations in 52 genes). This suggests that the mutations in the pathoadaptive genes have been positively selected in the host airways. In agreement with this, we found a significant mutational signature of positive selection of SNPs in pathoadaptive genes ($P = 0.007$; dN/dS > 5; 15 nonsynonymous SNPs and none syonymous SNPs) whereas SNPs in the remainder of genes showed a signature of negative selection (dN/dS = 0.7). Mutations in pathoadaptive genes includes nonsynonymous mutations in *aceF*, *algU*, *bifA*, *gyrA*, *gyrB*, *lasR*, *mexA*, *mexB*, *mexZ*, *morA*, *mucA*, *nfxB*, *pvdS*, and *retS*.

Furthermore, several other of the mutations accumulated in lineages IT01, IT02, IT03, IT05, and IT06 may as well confer a selective advantage. For example, the exact C to T *phuR* promoter mutation (PAO1 genome position 5289158) found in isolates F23, F40, and F41 has previously been shown to increase the expression of the *Pseudomonas* heme uptake (phu) system; an adaptive trait shown to improve the uptake of iron from host hemoglobin [19]. Also, loss-of-function mutations in *rpoN* have frequently been found to be typical markers of CF lung infection [10], and in this study we found an *rpoN*(11-12insCT) frameshift mutation in IT02 strains H24 and H25.

Finally, we identified an A2058G mutation and a C2611T (*Escherichia coli* numbering) mutation in the 23S rRNA gene in the IT02 lineage (isolates H24 and H25) and the IT06 lineage (isolates F23, F40, and F41), respectively. The exact mutations have previously been shown to confer macrolide resistance in *P. aeruginosa* [20]. Thus, we hypothesize that the mutations have been selected due the use of azithromycin in clinic (see Methods section for description of antibiotic therapy), and in accordance with this the mutations were not observed until after the start of oral azithromycin treatment of the respective patients.

Genetic relationship between IT01-IT06 lineages and other P. aeruginosa strains

Next, we sought to determine the genetic relationship between lineages IT01-IT06 and other genome sequenced strains of *P. aeruginosa*. We *de novo* assembled the genome of the earliest isolate(s) of each of the six clonal lineages and aligned the genomes to a panel of 60 genome sequences of other strains of *P. aeruginosa*. The panel consisted of eight completed genomes of *P. aeruginosa* reference strains (PAO1 [23], PA14 [24], LESB58 [25], PA7 [26], DK02 [21], 2192 and C3719 [27], PACS2), and 53 incomplete draft genomes of strains DK01-DK53, which we recently genome sequenced in a study of Danish CF patients [9].

Using the Harvest suite [28], we aligned the panel of genomes and constructed a phylogenetic tree based on 188,727 SNPs identified in the core genome of the aligned strains (Fig. 3). Clustering of genomes based on core genome SNPs confirmed previous findings that, except for a few outlier strains (e.g. PA7), strains group into two major phylogenetic clusters containing strains PAO1 and PA14, respectively (Fig. 3a) [29, 30]. We observed no evidence that the geographical linked strains IT01-IT06 grouped more closely in the phylogenetic tree. However, we noted that strains IT01-IT06 all belong to the phylogenetic cluster containing strain PAO1 (Fig. 3a).

Genetic distances between closely related isolates from different countries

We found that the IT02 and IT03 lineages were closely related to isolates of the DK26 and DK06 clone types, respectively (Fig. 3b), which have been isolated from Danish CF patients [9]. The close genetic relationship was also supported by multilocus sequence typing (MLST) analysis, showing the IT02/DK26 strains to be of sequence type ST-27 and the IT03/DK06 strains to be of sequence type ST-17.

To further investigate these two cases, we repeated the phylogenetic analysis, now only including the relevant isolates (Fig. 4). Note, since both clone types DK06 and DK26 were found in multiple patients in the study of the Danish patients, we included the earliest isolate from all the respective patients in our analysis (DK06 and DK26 isolates were from years 2006–2010 and 2011–2012, respectively). Interestingly, we found that more diversity was present between isolates from Denmark relative to the diversity present between isolates from Italy and Denmark (Fig. 4). For example, only 47 SNPs separated the Italian isolate H4 from isolate 294 sampled from a Danish CF patient (P21F4), whereas Danish isolates of the same clone type were different by up to 265 SNPs (isolates 294 and 34) (Table 2).

Host adaptation is associated with reduction in metabolic capacity

Finally, we sought to investigate the phenotypic changes that occurred during the course of infection. Reductions in metabolic capacity, for example caused by *rpoN* mutations, have previously been associated with host adaptation [5, 22], so we used phenotype microarrays (Biolog) to characterize the metabolic capacity of three strains from each of the lineages IT01, IT02, and IT06. The metabolic capacities of isolates of both lineages IT01 and IT02 decreased over time (19 and 10 years, respectively), whereas no change was observed for lineage IT06 over 4 years of infection (Fig. 5).

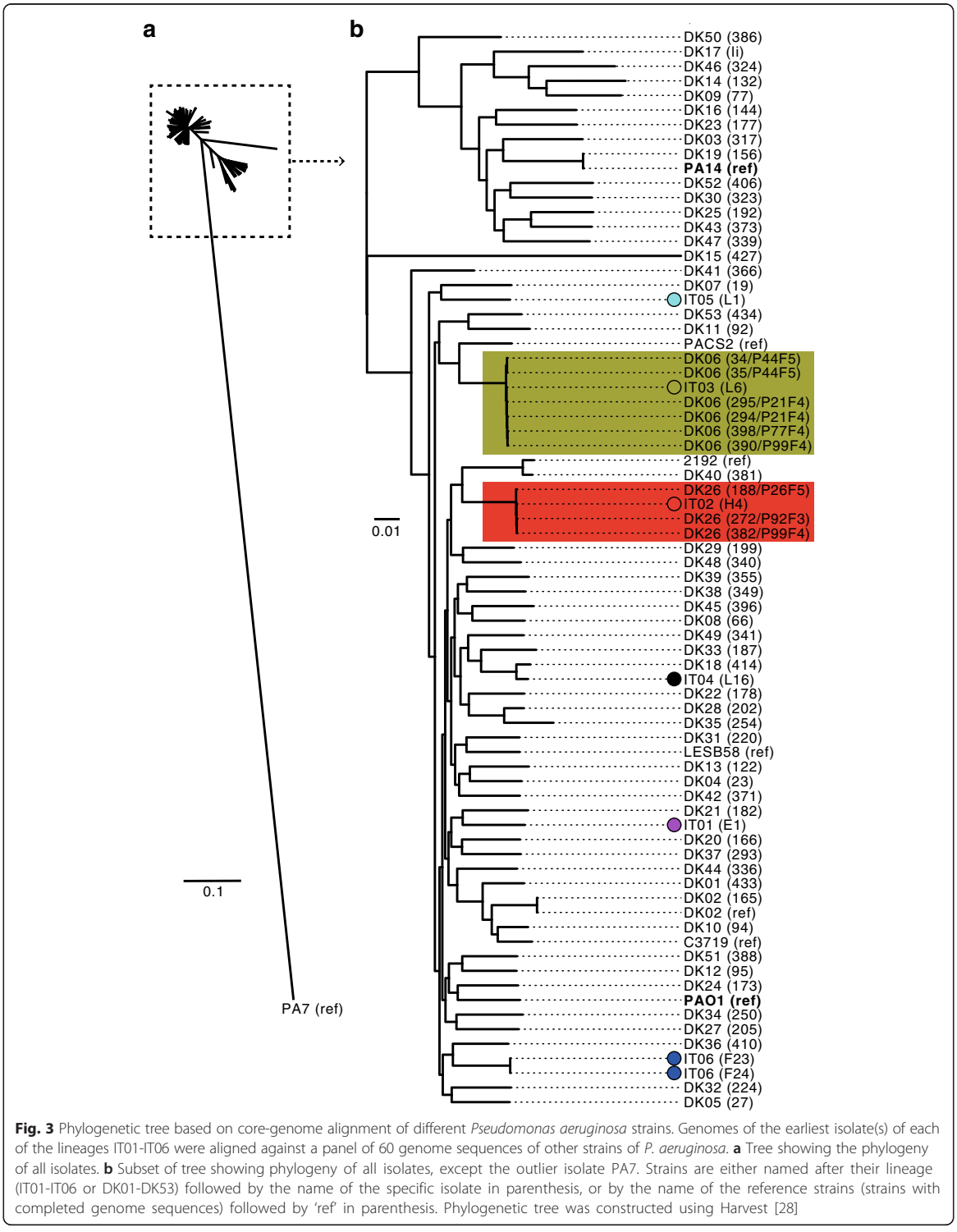


Fig. 3 Phylogenetic tree based on core-genome alignment of different *Pseudomonas aeruginosa* strains. Genomes of the earliest isolate(s) of each of the lineages IT01-IT06 were aligned against a panel of 60 genome sequences of other strains of *P. aeruginosa*. **a** Tree showing the phylogeny of all isolates. **b** Subset of tree showing phylogeny of all isolates, except the outlier isolate PA7. Strains are either named after their lineage (IT01-IT06 or DK01-DK53) followed by the name of the specific isolate in parenthesis, or by the name of the reference strains (strains with completed genome sequences) followed by 'ref' in parenthesis. Phylogenetic tree was constructed using Harvest [28]

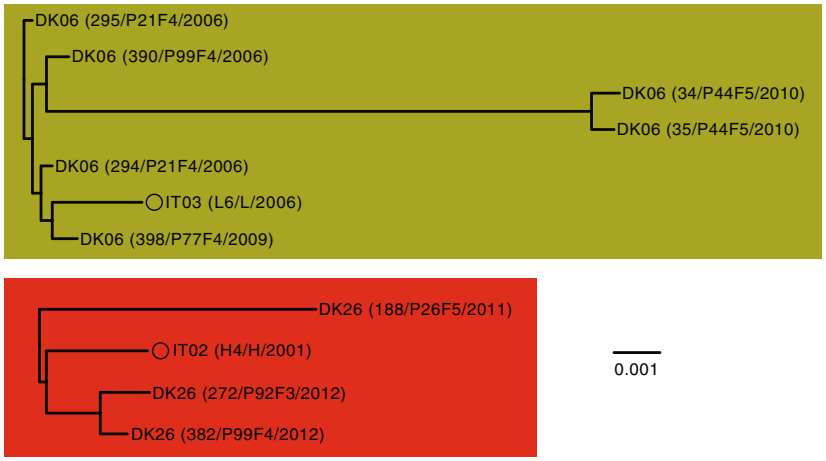


Fig. 4 Phylogenetic trees based on core-genome alignment of *Pseudomonas aeruginosa* strains of the IT02/DK26 and IT03/DK06 clonal complexes. Phylogenetic trees were constructed using Harvest [28]. Name, patient origin, and year of isolation are shown for each isolate in parenthesis

Mutants with loss-of-function mutations in the CbrAB two-component system is known to be unable or poor at utilizing a large range of carbon and nitrogen sources [31]. Therefore, we speculate that two nonsynonymous mutations accumulated in *cbrAB* may explain the loss of metabolic capacity in isolates H18 and H24 of the IT02 lineage.

In the IT01 lineage, several mutations may explain the reduced metabolic capacities of isolates E19 and E43 from years 2006 and 2012, respectively, relative to the ancestral isolate E1 from 1993. Both isolates E19 and E43 accumulated nonsynonymous mutations in genes *aceF*, *dgcB*, *gcdH*, *glcD*, *glnD*, and *vanA* belonging to PseudoCap function classes 'Amino acid biosynthesis and metabolism' and/or 'Carbon compound catabolism'. In addition, isolate E43 harbor mutations in genes *cysB*, *pdxJ*, PA4910, *sdhA* that are involved in biosynthesis of amino acids, biosynthesis of cofactors, amino acid transport, and central metabolism, respectively. We suggest that the mentioned mutations may explain the reduction of metabolic capacity in the IT01 lineage.

Changes in mucoidity, motility, ciprofloxacin susceptibility, and protease production

Besides reduction in metabolic capacity, the CF lung infection is associated with the appearance of a number of other phenotypes of which many are not usually observed among environmental, wild type isolates of this species. For example, mucoid colony formation, reduction in secretion of extracellular proteases, loss of flagella dependent motility, and antibiotic resistance

Table 2 Genetic distances (SNPs) between *Pseudomonas aeruginosa* strains

IT02/DK26	188/P26F5	272/P92F3	382/P99F4	H4			
188/P26F5	0	197	108	204			
272/P92F3	197	0	41	108			
382/P99F4	193	41	0	96			
H4	204	108	96	0			
IT03/DK06	294/P21F4	295/P21F4	34/P44F5	35/P44F5	398/P77F4	390/P99F4	L6
294/P21F4	0	18	265	259	21	26	47
295/P21F4	18	0	256	258	28	25	58
34/P44F5	265	256	0	23	259	262	277
35/P44F5	259	258	23	0	263	260	277
398/P77F4	21	28	259	263	0	26	53
390/P99F4	26	25	262	260	26	0	64
L6	47	58	277	277	53	64	0

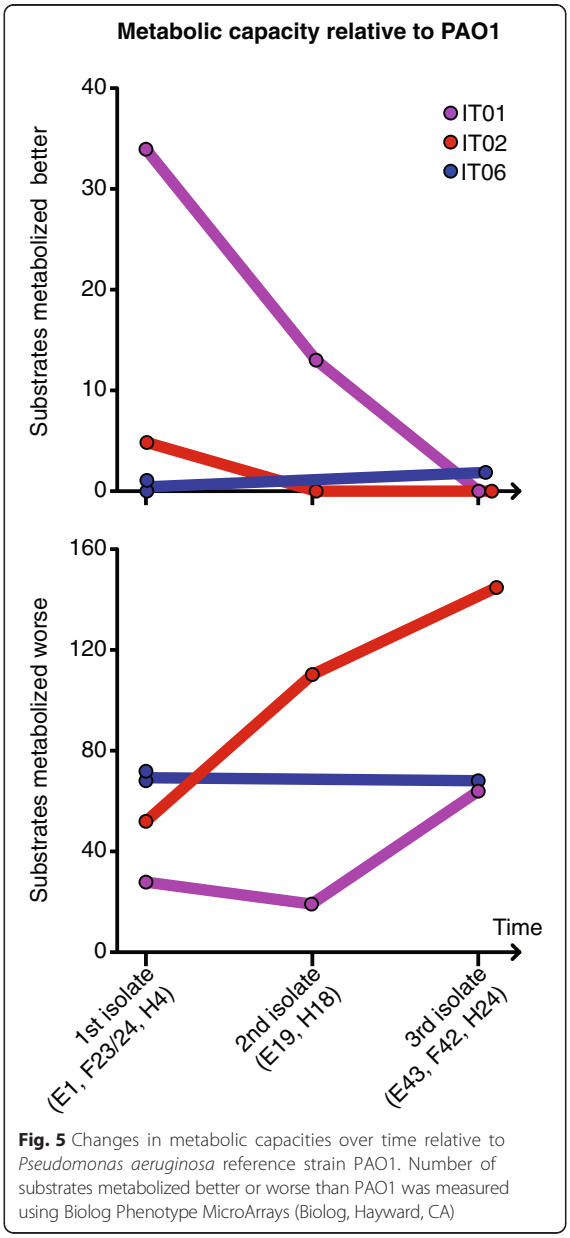


Fig. 5 Changes in metabolic capacities over time relative to *Pseudomonas aeruginosa* reference strain PAO1. Number of substrates metabolized better or worse than PAO1 was measured using Biolog Phenotype MicroArrays (Biolog, Hayward, CA)

are characteristics of isolates from chronic CF infections [32–34].

We phenotypically characterized the 25 isolates of clone types IT01-IT03 and IT05-06, and in accordance with previous studies, we found that nearly all isolates exhibited reduced swimming and swarming motility and protease secretion relative to the wild type-like reference strain PAO1 [35] (Fig. 2). The only isolates showing swimming and/or swarming capabilities better than PAO1 were L11, F4, and H1. Furthermore, three of the patients (E, F, and H) harbored mucoid isolates, and the two latest isolates of the IT02 lineage (H24 and H25) were resistant to ciprofloxacin.

The relative large numbers of mutations that appear in each of the clonal lineages make it difficult to associate specific mutations with phenotypic changes. Nonetheless, only mutations GacS(L309Q) and GacA(Y186stop), respectively, discriminated isolates F4 and H1 from the other isolates of the IT05 lineage. In agreement with literature [36, 37], this suggest that the mutations in the GacAS two-component regulatory system is the cause of increased motility and decreased extracellular protease activity of isolates F4 and H1. Furthermore, we suggest that a GyrA(T83I) mutation, which has previously been associated with ciprofloxacin resistance [38], explains ciprofloxacin resistance observed in isolates H24 and H25. Also, we suggest that mucoidity of isolates E37 and E42 may be caused by a nonsense mutation in *mucA* present in these two, but no other, isolates of the IT01 lineage.

In the IT02 lineage, a reduction of extracellular protease activity in isolates H18, H24, and H25 coincides with the fixation of a missense mutation in *lasR* encoding a regulator required for the expression of virulence-associated extracellular proteases LasA and LasB [39, 40]. In the same isolates, we suggest a loss of swimming motility to be caused by a missense mutation in *fleQ* encoding a positive regulator of flagella biosynthesis genes [41].

Discussion

We have gained insight into *P. aeruginosa* infections using a collection of isolates sampled over 19 years from four Italian CF patients. We sequenced the genomes of 26 of the isolates and correlated identified mutations to changes in relevant phenotypes and antibiotic treatment used in the clinic. While there are several other clinical collections of *P. aeruginosa* that have been genome sequenced [2], this is the first genome analysis of isolates from Italian CF patients. We therefore anticipate that our study may serve as a reference for future research and helps to obtain a more complete basis for understanding *P. aeruginosa* infections in CF patients.

We found the genomic evolution of lineages from Italian patients to resemble the evolution of *P. aeruginosa* lineages from other countries. This included similar observed rates of mutation and evidence for host-associated selection for mutations in pathoadaptive genes. Accordingly, our results support previous findings of significant mutational signature of positive selection in relatively few pathoadaptive genes in contrast to neutral change in the large remainder of genes. Inter-study parallelism of mutations in the same pathoadaptive genes may by part be driven by similarities in antibiotic treatment regimes. However, we also found parallelism of mutations in genes that

are not directly associated to antibiotic resistance (e.g. *aceF*, *bifA*, *morA*, and *retS*), and this may reflect similarities in host-dependent selective forces.

Unlike other studies [5, 8, 9, 42–44], we did not identify hypermutable lineages. Since our study only encompasses four patients, the lack of hypermutators may be accidental, and hypermutators may eventually appear as these are observed more frequently in late stage infections [42].

Patients were predominantly infected with strains unique to the particular patient. One exception was clone type IT05 which was shared by three patients. Despite of IT05's occurrence in multiple patients, the IT05 clone type resembled a wild type phenotype, and this may explain why other clone types with typical host-associated phenotypes replace IT05 in all patients. As such, the pattern of strain replacement may be due to initial infections from a strain that reside in a common environmental source to which the patients are frequently exposed, and that this strain is subsequently replaced by other more rare, but also more fit, strains.

While other studies have given insight into *P. aeruginosa* within-host population diversity [5, 7, 45, 46], further investigations are needed to make conclusions about the diversity of *P. aeruginosa* in our four patients. However, we conducted an inter-clonal genomic analysis to determine the genetic relationship between lineages IT01-IT06 and a panel of 60 genome sequenced strains of *P. aeruginosa* from other countries. Hereby, we showed that lineages IT02 and IT03 were closely related to lineages DK26 and DK06, respectively, which have been isolated from Danish CF patients [9]. Since more diversity was present between isolates from Denmark relative to the diversity present between isolates from Italy and Denmark, our findings show that one must be cautious of using genome analysis to infer the country origin of lineages.

Conclusions

This is the first whole-genome analysis of *P. aeruginosa* isolated from Italian CF patients, and together with both phenotypic and clinical information this dataset facilitates a better understanding of *P. aeruginosa* within-host genomic evolution, transmission, and population genomics. This may help the design of future intervention strategies for the clinical setting.

Methods

Bacterial isolates

The 35 *P. aeruginosa* isolates used in this study originates from the bacterial collection available at the Tuscan Regional Referral Center for Cystic Fibrosis in Florence, Italy. At this center, *P. aeruginosa* isolates sampled from CF patients have been stored twice a year since 1993. All patients enrolled in the study were in follow-up according to published guidelines [47, 48]. CF diagnosis was based on clinical features of the disease and concentration of chloride in sweat >60 mmol/liter [49]. Patients were regularly examined every 3 months. Data regarding their weight, height, body mass index (BMI), forced expiratory volume in one second (FEV_1), microbiological status (including antibodies against *P. aeruginosa*) and antibiotic treatments were stored in the database. Cough swabs or sputum samples were processed following the national and international guidelines [50] (https://www.cysticfibrosis.org.uk/media/82034/CD_Laboratory_Standards_Sep_10.pdf).

The patients gave written consent to participate and for publication of their details. The study and use of bacterial isolates has been approved by the local ethics committee at the Department of Paediatric Medicine, Anna Meyer Children's University Hospital, Florence, Italy (approval no. 210).

Antibiotic therapy principles in the clinic

Early eradication treatment was started at first detection of *P. aeruginosa* by administration of oral ciprofloxacin in combination with either nebulized colistin for three weeks or three months or inhaled tobramycin [48, 51, 52]. When *P. aeruginosa* reappeared after eradication therapy, treatment with oral ciprofloxacin and inhaled antibiotics was restarted.

At development of chronic *P. aeruginosa* infection [53], intravenous antibiotic treatment was administered together with inhaled colistin or inhaled tobramycin. Patients also received treatments with oral azitromycin [54]. During follow-up, in the case of mild pulmonary exacerbation, all patients were treated with oral ciprofloxacin in combination with inhaled antibiotics. Severe pulmonary exacerbations were treated with parenteral antibiotic treatments (ceftazidime or meropenem in combination with tobramycin once a day) according to suggested dosage [47].

Patient information

Patient E (cystic fibrosis transmembrane conductance regulator gene (*CFTR*) genotype F508del/F508del) had his first *P. aeruginosa* colonization at the age of 19 months and developed chronic infection at the age of three years. He remained stable without pulmonary exacerbation until 2012. Due to excellent lung function (FEV_1 = 115 % of predicted) he was treated only intermittently with oral ciprofloxacin in combination with inhaled antibiotics. Oral azithromycin was started in 2006.

Patient F (*CFTR* genotype 711 + 19 A/T /3272-9 A/t) had her first *P. aeruginosa* colonization at the age of 33 months and developed chronic infection at the age of four years. She had a stable FEV_1 (60 %) until 2008,

when she began a progressive clinical deterioration with FEV_1 (20 %) in 2012. She was frequently treated with high doses of ciprofloxacin with inhaled antibiotics from 2001 until 2010, with repeated parenteral antibiotic cycles with ceftazidime or meropenem in association with tobramycin once a day. Oral azithromycin was started in 2004.

Patient H (*CFTR* genotype F508del/F508del) had his first *P. aeruginosa* colonization at the age of 8 months and was chronically infected at the age of 1 year. After chronic *P. aeruginosa* infection was developed, he was treated with fluoroquinolones and tobramycin by inhalation. Oral quinolones were used to treat mild pulmonary exacerbations. He underwent parenteral antibiotic treatment (ceftazidime and tobramycin) twice from 2011 to 2013 for severe pulmonary exacerbations. His lung function declined from 2006 ($FEV_1 = 80$ %) to 2011 ($FEV_1 \leq 40$ %). Oral azithromycin was started in 2005. In addition, tetracyclines, trimethoprim-sulfamethoxazole, and linezolid were used to treat methicillin-resistant *Staphylococcus aureus*.

Patient L (*CFTR* genotype F508del/N1303K) had her first *P. aeruginosa* colonization at the age of 27 months. She developed chronic infection at the age of 20 years. However, her clinical conditions were stable and she had good lung function ($FEV_1 = 80$ %). After chronic *P. aeruginosa* lung infection developed, she was treated with antibiotics by inhalation. She used oral quinolones for mild pulmonary exacerbations together with parenteral antibiotic treatment (ceftazidime and tobramycin), which was necessary twice in the period 2011 to 2013. She did not tolerate treatment with oral macrolides.

ArrayTube genotyping

P. aeruginosa genotypes were determined using an ArrayTube multimarker microarray targeting 13 SNPs in the core genome and additional genetic markers in the accessory genome (Clondiag Chip Technologies, Germany) [13]. The ArrayTube genotyping assay was performed according to the protocol provided by the manufacturer.

Genome sequencing

Genomic DNA was prepared from *P. aeruginosa* isolates on a QIAcube system using a DNeasy Blood and Tissue Kit (Qiagen). Genomes were sequenced by BGI (Shenzhen, China) to an average coverage depth of at least 75-fold (range 75- to 139-fold) on an Illumina HiSeq2000 platform generating 100-nt paired-end reads.

Mutation detection and construction of phylogenetic trees

Mutations that had accumulated in *P. aeruginosa* lineages IT01-IT03 and IT05-IT06, were identified as previously described (without any modifications) [9]. Based on the identified SNPs, we computed maximum-parsimonious phylogenetic trees using PAUP* version 4.0b10 [55]. Consistency indexes were calculated as the number of mutations divided by the minimum number of mutational events required to explain phylogenies. The consistency index will equal one when there is no homoplasy.

Estimation of mutation rates

Bayesian analysis of evolutionary rate was performed using BEAST, version 1.7.0 [56], with a lognormal relaxed molecular clock model and a general time-reversible substitution model. Mutation rate of the IT01 lineage was calculated from a chain length of 50 million steps, sampled every 5,000 steps. The first 5 million steps were discarded as a burn-in. The ESS of all parameters were >1,000 as calculated by Tracer, version 1.5 (available from http://beast.bio.ed.ac.uk/Tracer), which was also used to calculate the 95 % HPD confidence intervals of the mutation rate (i.e. an interval within which the modeled parameter resides with 95 % probability). Mutation rates of the other lineages (IT02, IT03, IT05, IT06) could not be estimated since the number and temporal distribution of isolates were insufficient to obtain proper estimates of the mutation rates (ESS of modeled parameters < 10).

De novo assembly and inter-clonal whole-genome alignments

We *de novo* assembled the genome of the earliest isolate(s) of each of the six clonal lineages (E1, F23, F24, H4, L1, L6, L16) and genomes of representatives of strains DK01-DK53, which we recently genome sequenced in a study of Danish CF patients [9].

Sequence reads from each isolate were error corrected using ALLPATHS-LG's stand-alone error correction tool [57] and *de novo* assembled using the de Bruijn graph-based assembler Velvet (version 1.2.08) [58]. For each sample several assemblies were run. This implies that for each dataset 'velveth' command was executed using *k*-mer sizes in the range of 35 to 95. Next, the 'velvetg' command was run using the parameters: min_contig_lgth = 400, exp_cov = auto, and scaffolding = no. Based on the number of contigs, the best cumulative rank for N50, and the length of the largest contig, the best *k*-mer size was selected and the exp_cov was noted. A final assembly was performed using the best *k*-mer size for velveth and with velvetg using the parameters: min_contig_lgth = 400, scaffolding = no, and the exp_cov that was calculated in the first ensemble of assemblies. The size range of the *de novo* assembled genomes of strains of lineages IT01-06 was 6.3-6.9 Mbp, and each genome shared 95.8–97.0 % of the content *P. aeruginosa* reference strain PAO1.

De novo assembled genomes and completed genomes from the public domain [59] were aligned against each other using the Harvest suite [28] with default parameters using all genomes (option '–c' set to 'YES') and with genome sequence of strain PAO1 as reference.

MLST sequence types were determined using a public available online tool developed for use on *de novo* assembled genomes [60]. Sequence types of lineages IT01-06 were ST-111, ST-27, ST-17, ST-389, ST-1748, and ST-348, respectively.

Biolog phenotype profiling

Phenotype MicroArray (Biolog, Hayward, CA) experiments were performed in duplicate according to the manufacturer's instructions [61, 62]. *P. aeruginosa* strains were streaked on LB agar plates and incubated at 37 °C until colonies appeared on the plates (16–30 hours (h)). Cells were swabbed from the plates and suspended in IF-0 GN Base (inoculation fluid) at a density corresponding to 42 % transmittance in the Biolog turbidimeter. The cell suspensions were diluted 1:6 in IF-0 minimal medium containing Biolog redox dye mixture D (tetrazolium), and 100-μL aliquots were added to carbon-source plates (PM1 and PM2A). For the nitrogen-source plate (PM3B), inoculations were supplemented with 30 mM glucose and 2 μM ferric citrate. The plates were incubated at 37 °C in an OmniLog plate reader (Biolog) for 72 h, and growth/respiration was measured kinetically by determining the colorimetric reduction of tetrazolium dye. Export of OmniLog data was performed using OmniLog OL_FM/Kin 1.20.02 software (Biolog). The average area beneath each kinetic curve was used for analysis. OD600 was measured after 60 h of growth at 37 °C, and relative catabolic capacities of individual substrates were determined as previously described [63].

Motility assays

The motility of all isolates was tested on ABT minimal medium supplemented with 0.5 % Casamino acids and 0.5 % glucose as previously described [64]. Briefly, agar plates were inoculated with single colonies on the top or in the middle of the plates to measure swarming or swimming motility, respectively. Swimming and swarming were measured on 0.3 % (wt/vol) and 0.6 % (wt/vol) agar, respectively, with incubation for 24 h at 30 °C. The motility zones in triplicates were measured relative to the motility of the reference strain PAO1.

Skim milk protease assay

The production of protease was determined by applying one colony of the isolate grown on Luria-Bertani (LB) medium agar plates to skim milk agar plates (LB agar with 10 % skim milk). The plates were incubated for 24 h at 37 °C, and the clearing zones were measured.

Ciprofloxacin susceptibility testing

Ciprofloxacin susceptibility was tested by the Bauer-Kirby agar disk diffusion method. Discs (Neo-Sensitabs) were purchased from Rosco Diagnostica (Taastrup, Denmark). The diameters of clearing zones were measured in millimeters (mm) after 20 h at 37 °C.

Availability of data and materials

Sequence reads from all *P. aeruginosa* isolates are deposited in the Short Read Archive under accession number PRJEB8020 [EMBL:PRJEB8020].

Abbreviations

BMI: Body mass index; CF: Cystic fibrosis; CFTR: Cystic fibrosis transmembrane conductance regulator gene; dN: Rate of nonsynonymous genetic changes; dS: Rate of synonymous genetic changes; ESS: Effective sample size; FEV_1: Forced expiratory volume in one second; HPD: Highest posterior density; h: Hour; mm: Milimeter; MLST: Multilocus sequence typing; Phu: *Pseudomonas* heme utilization; SNP: Single nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SM, GT, and HKJ jointly supervised and conceived the study. RLM, DD, LMS, OC, SM, GT, and HKJ designed the experiments. DD, SC, and TC collected clinical samples and provided clinical information. DD and LMS carried out experiments in the laboratory. BP *de novo* assembled genomes. RLM and LMS conducted whole-genome sequence analysis. RLM, DD, LMS, OC, SM, GT, and HKJ analyzed and interpreted the results. RLM wrote the manuscript. DD, LMS, BP, OC, SC, SM, GT, and HKJ helped write the manuscript and provided revisions.

Acknowledgments

We thank the Director of the CF Centre of Florence Dr. Cesare Braggion for discussion and supporting the research. The Brunelleschi Rotary Club Florence supported DD travel costs and stipend, and the Novo Nordisk Foundation supported HKJ as a clinical research stipend.

Author details

[1]Department of Immunology and Microbiology, Costerton Biofilm Center, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [2]Center for Genomic Medicine, Rigshospitalet, Copenhagen, Denmark. [3]Department of Paediatric Medicine, Cystic Fibrosis Centre, Anna Meyer Children's University Hospital, Florence, Italy. [4]Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark. [5]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. [6]Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark.

Received: 8 April 2015 Accepted: 12 October 2015
Published online: 19 October 2015

References

1. Wilson DJ. Insights from genomics into bacterial pathogen populations. PLoS Pathog. 2012;8, e1002874.
2. Marvig RL, Sommer LM, Jelsbak L, Molin S, Johansen HK. Evolutionary insight from whole-genome sequencing of Pseudomonas aeruginosa from cystic fibrosis patients. Future Microbiol. 2015;10.
3. Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF, Tummler B. Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs. Environ Microbiol. 2011;13:1690–704.

4. Dettman JR, Rodrigue N, Aaron SD, Kassen R. Evolutionary genomics of epidemic and nonepidemic strains of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A. 2013;110:21065–70.
5. Feliziani S, Marvig RL, Lujan AM, Moyano AJ, Di Rienzo JA, Krogh Johansen H, et al. Coexistence and Within-Host Evolution of Diversified Lineages of Hypermutable Pseudomonas aeruginosa in Long-term Cystic Fibrosis Infections. PLoS Genet. 2014;10, e1004651.
6. Jeukens J, Boyle B, Kukavica-Ibrulj I, Ouellet MM, Aaron SD, Charette SJ, et al. Comparative genomics of isolates of a Pseudomonas aeruginosa epidemic strain associated with chronic lung infections of cystic fibrosis patients. PLoS One. 2014;9, e87611.
7. Markussen T, Marvig RL, Gomez-Lozano M, Aanaes K, Burleigh AE, Hoiby N, et al. Environmental heterogeneity drives within-host diversification and evolution of Pseudomonas aeruginosa. MBio. 2014;5:e01592–01514.
8. Marvig RL, Johansen HK, Molin S, Jelsbak L. Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. PLoS Genet. 2013;9, e1003741.
9. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat Genet. 2015;47:57–64.
10. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, et al. Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. Proc Natl Acad Sci U S A. 2006;103:8487–92.
11. Burns JL, Gibson RL, McNamara S, Yim D, Emerson J, Rosenfeld M, et al. Longitudinal assessment of Pseudomonas aeruginosa in young children with cystic fibrosis. J Infect Dis. 2001;183:444–52.
12. Cramer N, Wiehlmann L, Ciofu O, Tamm S, Hoiby N, Tummler B. Molecular epidemiology of chronic Pseudomonas aeruginosa airway infections in cystic fibrosis. PLoS One. 2012;7, e50731.
13. Wiehlmann L, Wagner G, Cramer N, Siebert B, Gudowius P, Morales G, et al. Population structure of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A. 2007;104:8101–6.
14. Johansen HK, Aanaes K, Pressler T, Nielsen KG, Fisker J, Skov M, et al. Colonisation and infection of the paranasal sinuses in cystic fibrosis patients is accompanied by a reduced PMN response. J Cyst Fibros. 2012;11:525–31.
15. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 2000;15:496–503.
16. Snyder LA, Loman NJ, Faraj LA, Levi K, Weinstock G, Boswell TC, et al. Epidemiological investigation of Pseudomonas aeruginosa isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing. Euro Surveill. 2013;18.
17. Damkiaer S, Yang L, Molin S, Jelsbak L. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. Proc Natl Acad Sci U S A. 2013;110:7766–71.
18. Klockgether J, Miethke N, Kubesch P, Bohn YS, Brockhausen I, Cramer N, et al. Intraclonal diversity of the Pseudomonas aeruginosa cystic fibrosis airway isolates TBCF10839 and TBCF121838: distinct signatures of transcriptome, proteome, metabolome, adherence and pathogenicity despite an almost identical genome sequence. Environ Microbiol. 2012;15(1):191–210.
19. Marvig RL, Damkiaer S, Khademi SM, Markussen TM, Molin S, Jelsbak L. Within-Host Evolution of Pseudomonas aeruginosa Reveals Adaptation toward Iron Acquisition from Hemoglobin. MBio. 2014;5.
20. Marvig RL, Sondergaard MS, Damkiaer S, Hoiby N, Johansen HK, Molin S, et al. Mutations in 23S rRNA confer resistance against azithromycin in Pseudomonas aeruginosa. Antimicrob Agents Chemother. 2012;56(8):4519–21.
21. Rau MH, Marvig RL, Ehrlich GD, Molin S, Jelsbak L. Deletion and acquisition of genomic content during early stage adaptation of Pseudomonas aeruginosa to a human host environment. Environ Microbiol. 2012;14:2200–11.
22. Yang L, Jelsbak L, Marvig RL, Damkiaer S, Workman CT, Rau MH, et al. Evolutionary dynamics of bacteria in a human host environment. Proc Natl Acad Sci U S A. 2011;108:7481–6.
23. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, et al. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. Nature. 2000;406:959–64.
24. Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S, et al. Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. Genome Biol. 2006;7:R90.
25. Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F, et al. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. Genome Res. 2009;19:12–23.
26. Roy PH, Tetu SG, Larouche A, Elbourne L, Tremblay S, Ren Q, et al. Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. PLoS One. 2010;5, e8842.
27. Mathee K, Narasimhan G, Valdes C, Qiu X, Matewish JM, Koehrsen M, et al. Dynamics of Pseudomonas aeruginosa genome evolution. Proc Natl Acad Sci U S A. 2008;105:3100–5.
28. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15:524.
29. Hilker R, Munder A, Klockgether J, Losada PM, Chouvarine P, Cramer N, et al. Interclonal gradient of virulence in the Pseudomonas aeruginosa pangenome from disease and environment. Environ Microbiol. 2015;17(1):26–46.
30. Stewart L, Ford A, Sangal V, Jeukens J, Boyle B, Kukavica-Ibrulj I, et al. Draft genomes of 12 host-adapted and environmental isolates of Pseudomonas aeruginosa and their positions in the core genome phylogeny. Pathog Dis. 2014;71:20–5.
31. Nishijyo T, Haas D, Itoh Y. The CbrA-CbrB two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in Pseudomonas aeruginosa. Mol Microbiol. 2001;40:917–31.
32. Doggett RG, Harrison GM, Wallis ES. Comparison of Some Properties of Pseudomonas Aeruginosa Isolated from Infections in Persons with and without Cystic Fibrosis. J Bacteriol. 1964;87:427–31.
33. Luzar MA, Montie TC. Avirulence and altered physiological properties of cystic fibrosis strains of Pseudomonas aeruginosa. Infect Immun. 1985;50:572–6.
34. Mahenthiralingam E, Campbell ME, Speert DP. Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis. Infect Immun. 1994;62:596–605.
35. Holloway BW, Krishnapillai V, Morgan AF. Chromosomal genetics of Pseudomonas. Microbiol Rev. 1979;43:73–102.
36. Goodman AL, Kulasekara B, Rietsch A, Boyd D, Smith RS, Lory S. A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in Pseudomonas aeruginosa. Dev Cell. 2004;7:745–54.
37. Kay E, Humair B, Denervaud V, Riedel K, Spahr S, Eberl L, et al. Two GacA-dependent small RNAs modulate the quorum-sensing response in Pseudomonas aeruginosa. J Bacteriol. 2006;188:6026–33.
38. Cabot G, Ocampo-Sosa AA, Dominguez MA, Gago JF, Juan C, Tubau F, et al. Genetic markers of widespread extensively drug-resistant Pseudomonas aeruginosa high-risk clones. Antimicrob Agents Chemother. 2012;56:6349–57.
39. Gambello MJ, Iglewski BH. Cloning and characterization of the Pseudomonas aeruginosa lasR gene, a transcriptional activator of elastase expression. J Bacteriol. 1991;173:3000–9.
40. Toder DS, Gambello MJ, Iglewski BH. Pseudomonas aeruginosa LasA: a second elastase under the transcriptional control of lasR. Mol Microbiol. 1991;5:2003–10.
41. Arora SK, Ritchings BW, Almira EC, Lory S, Ramphal R. A transcriptional activator, FleQ, regulates mucin adhesion and flagellar gene expression in Pseudomonas aeruginosa in a cascade manner. J Bacteriol. 1997;179:5574–81.
42. Ciofu O, Riis B, Pressler T, Poulsen HE, Hoiby N. Occurrence of hypermutable Pseudomonas aeruginosa in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. Antimicrob Agents Chemother. 2005;49:2276–82.
43. Oliver A, Canton R, Campo P, Baquero F, Blazquez J. High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. Science. 2000;288:1251–4.
44. Waine DJ, Honeybourne D, Smith EG, Whitehouse JL, Dowson CG. Association between hypermutator phenotype, clinical variables, mucoid phenotype, and antimicrobial resistance in Pseudomonas aeruginosa. J Clin Microbiol. 2008;46:3491–3.
45. Chung JC, Becq J, Fraser L, Schulz-Trieglaff O, Bond NJ, Foweraker J, et al. Genomic Variation among Contemporary Pseudomonas aeruginosa Isolates from Chronically Infected Cystic Fibrosis Patients. J Bacteriol. 2012;194:4857–66.
46. Williams D, Evans B, Haldenby S, Walshaw MJ, Brockhurst MA, Winstanley C, et al. Divergent, Coexisting, Pseudomonas aeruginosa Lineages in Chronic Cystic Fibrosis Lung Infections. Am J Respir Crit Care Med. 2015;191(7):775–85.
47. Doring G, Hoiby N. Early intervention and prevention of lung disease in cystic fibrosis: a European consensus. J Cyst Fibros. 2004;3:67–91.
48. Taccetti G, Bianchini E, Cariani L, Buzzetti R, Costantini D, Trevisan F, et al. Early antibiotic treatment for Pseudomonas aeruginosa eradication in

patients with cystic fibrosis: a randomised multicentre study comparing two different protocols. Thorax. 2012;67:853–9.
49. Farrell PM, Rosenstein BJ, White TB, Accurso FJ, Castellani C, Cutting GR, et al. Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. J Pediatr. 2008;153:S4–S14.
50. Saiman L, Siegel J. Infection control recommendations for patients with cystic fibrosis: microbiology, important pathogens, and infection control practices to prevent patient-to-patient transmission. Infect Control Hosp Epidemiol. 2003;24:S6–52.
51. Frederiksen B, Koch C, Hoiby N. Antibiotic treatment of initial colonization with Pseudomonas aeruginosa postpones chronic infection and prevents deterioration of pulmonary function in cystic fibrosis. Pediatr Pulmonol. 1997;23:330–5.
52. Valerius NH, Koch C, Hoiby N. Prevention of chronic Pseudomonas aeruginosa colonisation in cystic fibrosis by early treatment. Lancet. 1991;338:725–6.
53. Lee TW, Brownlee KG, Conway SP, Denton M, Littlewood JM. Evaluation of a new definition for chronic Pseudomonas aeruginosa infection in cystic fibrosis patients. J Cyst Fibros. 2003;2:29–34.
54. Saiman L, Marshall BC, Mayer-Hamblett N, Burns JL, Quittner AL, Cibene DA, et al. Azithromycin in patients with cystic fibrosis chronically infected with Pseudomonas aeruginosa: a randomized controlled trial. JAMA. 2003;290:1749–56.
55. Swofford DL. PAUP*, Phylogenetic Analysis Using Parsimony (PAUP*and Other Methods). Version 4. Sunderland, Massachusetts: Sinauer Associates; 2003.
56. Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol.
57. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108:1513–8.
58. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.
59. Winsor GL, Lam DK, Fleming L, Lo R, Whiteside MD, Yu NY, et al. Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. Nucleic Acids Res. 2011;39:D596–600.
60. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol. 2012;50:1355–61.
61. Bochner BR. New technologies to assess genotype-phenotype relationships. Nat Rev Genet. 2003;4:309–14.
62. Bochner BR. Global phenotypic characterization of bacteria. FEMS Microbiol Rev. 2009;33:191–205.
63. Cooper VS, Lenski RE. The population genetics of ecological specialization in evolving Escherichia coli populations. Nature. 2000;407:736–9.
64. Hentzer M, Riedel K, Rasmussen TB, Heydorn A, Andersen JB, Parsek MR, et al. Inhibition of quorum sensing in Pseudomonas aeruginosa biofilm bacteria by a halogenated furanone compound. Microbiology. 2002;148:87–102.

4.4 study 4:

Evolutionary and adaptational differences of *Pseudomonas aeruginosa* in primary ciliary dyskinesia and cystic fibrosis patients.

Sommer L. M.*, Alanin M. C.*, Marvig R. L., Nielsen K. G., Høiby N., von Buchwald C., Molin S., and Johansen H. K. (2015), *Manuscript in preparation*

Evolutionary and adaptational differences of *Pseudomonas aeruginosa* in primary ciliary dyskinesia and cystic fibrosis patients

Lea M. Sommer[1]*, Mikkel C. Alanin[2]*, Rasmus L. Marvig[4], Kim G. Nielsen[3], Niels Høiby[4,5], Christian von Buchwald[2], Søren Molin[1], Helle K. Johansen[1,4]

[1]*Novo Nordisk Foundation Center for Biosustainability, The Technical University of Denmark, Hørsholm, Denmark.*
[2]*Department of Otolaryngology - Head and Neck Surgery and Audiology, Copenhagen University Hospital, Rigshospitalet, Denmark*
[3]*Danish PCD Centre, Pediatric Pulmonary Service, Department of Paediatrics and Adolescent Medicine, Copenhagen University Hospital, Rigshospitalet, Denmark*
[4]*Department of Clinical Microbiology, Copenhagen University Hospital, Rigshospitalet, Denmark*
[5]*Institute of Immunology and Microbiology, University of Copenhagen, Denmark*

Corresponding author: Helle K. Johansen
* These two authors contributed equally to the study. Shared first authorship

Running title: Adaptation and evolution of *P. aeruginosa* in PCD and CF patients

ABSTRACT

Opportunistic infections by *Pseudomonas aeruginosa* is a major cause of morbidity and mortality in immune-compromised patients. Genome sequencing of *P. aeruginosa* isolates from cystic fibrosis patients has shown that persistence of clonal lineages of *P. aeruginosa* in the airways is facilitated by genetic adaptation. While these findings are of clinical importance, it is unknown whether they also apply to *P. aeruginosa* airway infections in patients with primary ciliary dyskinesia (PCD). Both diseases are caused by disruption of the mucociliary clearance; however, while disruption of mucociliary clearance in PCD is caused by unfunctional ciliary, the failure is caused by thickening of the mucus in CF. Consequently, we hypothesize that investigating parallelism in within host evolution across opportunistic infections across diseases can infer similarities in the selective pressures governing infection in PCD and CF, respectively, and help to separate the role of unfunctional ciliary versus thick mucus in relation to selective pressures.

Here, we genome sequenced 35 longitudinally collected *P. aeruginosa* isolates and characterised 41 isolates phenotypically, from 12 patients with PCD. This comprised on average 2.9 (2-5) sequenced isolates per patient with an average timespan of 2 years (0.4-3.6), and phenotypical analysis of 3.4 isolates per patient (2-6) with an average timespan of 2.4 years (0.4-4.8).

We identified 14 clone types, and 9 out of 12 patients were infected with a unique clone type. Among these clone types we found eight genes to be mutated in parallel, suggesting that mutations were positively selected. Further, six of these genes have been found to be important for infections in CF airways. There was not much overlap of the bacterial phenotypic adaptations between the two diseases. In opposition to what is described in CF, PCD isolates show either no change or increased ability to produce biofilm over time, and a general preservation of swimming motility.

PCD and CF patients have comparable airway disease but PCD patients have in general a much milder degree of airway infections with better prognosis than CF patients. This study sheds light on the differences between genotype and phenotype adaptation of *P. aeruginosa* to the airways of PCD patients, as compared to CF patients. Our findings may resolve the basic questions of evolutionary adaptation of *P. aeruginosa* in airway infections in PCD and CF, and through this knowledge we provide important information that may consider new treatment strategies for *P. aeruginosa* infections.

INTRODUCTION

Pseudomonas aeruginosa is an opportunistic pathogen that frequently causes chronic infections in the upper and lower airways of both PCD and CF patients and it is the primary cause of morbidity and mortality in CF patients [1-3]. Patients are usually intermittently colonised in childhood, but despite intensive antibiotic therapy a chronic infection is rarely prevented. Thus, approximately 30% of all Danish CF patients are chronically infected in early adulthood [4], and approximately the same percentage of PCD patients eventually become chronically infected [5].

Investigations of *P. aeruginosa* have shown how the clone types can adapt and evolve to the CF lung environment, via phenotypic and mutational changes [6-13]. The most important characteristics include overproduction of alginate (mucoidity), slow growth, altered biofilm mode of growth, loss of motility, quorum sensing, and protease production [6,9,14-17]. These adaptational changes have to a large degree been linked with antibiotic treatments, the activity of the immune defence, and specific physical characteristics of the CF mucus. However it is not clear to what degree these factors independently and together impact the adaptation and evolution of *P. aeruginosa.*

These adaptive and evolutionary investigations have to our knowledge not been described in PCD patients and for the first time we describe the genetic and phenotypic evolution of distinct clonal lineages of *P. aeruginosa* from the PCD airways, and compare these with the evolutionary processes previously identified in the airways of CF patients [6,8,14-16,18,19].

Stasis of respiratory secretions is inevitable in patients with primary ciliary dyskinesia (PCD) and cystic fibrosis (CF) predisposing both patient groups to airway infections, inflammations and declining lung function. Inhaled pathogens are trapped in mucus in the airways as a first line of defence. In non-diseased airways, these entrapped pathogens are then removed by mucociliary clearance: Constant beating of motile cilia drives the mucus and pathogens up and out of the airways, where it is cleared by expectoration or ingestion. However, in PCD and CF this function of airway clearance is impaired: either directly by mutations resulting in structural or functional abnormalities of the cilia [20], or indirectly by mutations in a chloride ion channel resulting in thickened mucus in the airways and thereby impairing the ciliary movements [21].

We hypothesise, that if the main selective pressures found in the CF airways are the host immune defence system and the constant presence of antibiotics, then the adaptation and evolution of *P. aeruginosa* should not differ significantly between the two groups of patients. However, if the main selective pressure is the immediate environment comprising the structure and composition of the mucus found in the airways, then the adaptation and evolutionary pathways should differ significantly. The information provided by this study is of paramount importance to resolve the basic questions of evolutionary drive in *P. aeruginosa* in airway infections in both PCD and CF, and this knowledge may provide new insight and considerations that may help form new effective treatment strategies for infections with *P. aeruginosa*.

MATERIALS AND METHODS

Patients

We included 12 patients with definitive PCD diagnosis based on presentation of the characteristic clinical phenotype, ciliary ultrastructural defects visualized by electron microscopy, high speed video recordings showing abnormal ciliary function and/or a genetic mutation recognized to cause PCD [22]. All patients were diagnosed and treated at the Danish PCD Centre, at Rigshospitalet in Copenhagen. Patients are followed on a routine basis every three months, where a sputum

sample is collected, and clinical data such as medication, height, weight, and lung function are recorded.

Antibiotic treatment of patients

Chronic lung infection with *P. aeruginosa* is treated the same way in PCD and CF patients [5,23]. Treatment is initiated when a sample from the lower airways is positive for a relevant pathogen, including instances where clinical symptoms are absent. Chronically infected patients are treated with a combination of intravenous antibiotic therapy according to susceptibility testing either with a broad-spectrum beta-lactam or a carbapenem together with an aminoglycoside or a quinolone, every third month.

Clinical P. aeruginosa isolates

Isolation and identification of *P. aeruginosa* from sputum samples was carried out as previously described, independent of underlying condition. In brief, sputum samples were cultured at 37°C on standard agar media for 2 to 5 days. The media included a Sabouraud plate, a 7% NaCl plate, a "blue plate" (modified Conradi Drigalski's medium) selective for Gram-negative rods and non-selective media including 5% Danish blood agar and chocolate agar [24]. Selected isolates were stored from PCD patients with chronic *P. aeruginosa* infection. Chronic infection status was based on *P. aeruginosa* positive airway samples in combination with elevated serum precipitating antibodies [5].

Genome sequencing and analysis of PCD isolates

Preparation of Genomic DNA from 35 longitudinally collected *P. aeruginosa* isolates from 12 PCD patients and the sequence analysis, including clone type identification, was performed as previously described [13].

The identification of significantly mutated genes was carried out using the same principles as in Marvig et al. [13]. All genes were considered separately dependent on their specific size incorporating the random roulette theory of random mutation generation. We further limited the number of significantly mutated genes by only including genes with a tenfold enrichment as compared to the generated number of mutated clones. Also, genes had to be mutated in more than two clone types, or mutated in exactly two clone types if we did not observe mutation in more than one clone type as generated by the 1000 iterations.

Growth rate analysis

All isolates were picked from freezing stock and incubated directly in a 96 well plate with 100µl Luria Bertani (LB) broth, leaving the outer wells as blank. The plates were incubated at 37°C in a microtitter plate reader (Holm & Halby, BioTek Instruments Inc., DK-2605, model: ELX808IU) with OD (600nm) readings every 20 minutes until all isolates had reached stationary phase. The growth rates were calculated from the exponential part of the growth curves, and all isolates were assayed with a minimum of three replicates and a maximum of seven.

Motility assays

Motility assays were performed on soft agar plates of LB medium (Twitching: 1% agar, Swimming: 0.3% agar). Plates were inoculated from single colonies using sterile toothpicks either by positioning the colony at the bottom of the plate (twitching) or in the middle of the agar (swimming), both were incubated 24 hours at 37°C. The swimming assay was carried out in 96 well microtitter plates. A positive well was indicated by a turbid well, whereas a negative was not turbid. All isolates were assayed with a minimum of three replicates and a maximum of four.

Protease assay

Secreted protease production was assayed using LB agar (1% agar) and 3% (w/vol) skimmed milk. 200µl agar was added to each well of a 96 well plate and after drying, a hole was made in each well, where 10µl supernatant of an overnight liquid culture (also grown in LB at 37°C) was deposited. After 24h, wells were read as positive if the agar had become clear and negative if not. All isolates were assayed with a minimum of five replicates and a maximum of eight.

Attachment assay (Biofilm)

An overnight culture (also grown in LB at 37°C) was diluted 1:100 wherefrom 150µl was added to each well in a 96 well plate and a start OD (600nm) was measured. The plates were incubated at 37°C, 150rpm for 24h and an end point OD (600nm) was measured. The culture was removed and the plates washed three times in tab water, where after 200µl 0.01% crystal violet was added to each well. After 20 min at room temperature (RT) the crystal violet was removed and the plates was washed three times in tab water. 200µl 96% ethanol was added to the wells and left for 20 min at 240rpm at RT, where after the OD (620nm) was measured. For all plates, none of the outer wells were used for isolate growth, and all plates had six technical replicates of each isolate and six blanks (not counting the outer wells). All isolates were assayed with a minimum of three biological replicates and a maximum of seven.

Accession codes

Sequence reads from all *P. aeruginosa* isolates have been deposited in the Sequence Read Archive (SRA) under accession: Not deposited yet.

RESULTS

PCD isolate collection

In this study we have investigated 41 *P. aeruginosa* isolates longitudinally collected from 12 PCD patients with chronic lung infections [15,16]. The median age at the PCD diagnosis was 7.4 years (range: 0.0 - 29.4). The median age of the patients at the time of diagnosis of chronic lung infection was 15.5 years (range 9 – 59 years), and the median age at which we analysed the first *P. aeruginosa* isolate was 17.5 years (range 10- 62 years). On average the patients included in this study had the first *P. aeruginosa* cultured at 12 years of age (range 4 - 45).

To gain insight into the adaptation of *P. aeruginosa* to the airways of PCD patients, 35 longitudinal isolates were whole genome sequenced and 41 longitudinal isolates were phenotypically characterised. On average, we genome sequenced 2.9 isolates (range 2-5) per patient with an average timespan of 2.0 years (range 0.4-3.6) (Figure 1A), and phenotypically analysed 3.4 isolates (range 2-6) per patient with an average timespan of 2.4 years (range 0.4-4.8) (Figure 1B).

Phylogeny of the PCD isolates

From the genomic analysis of the 35 isolates we were able to identify 14 distinct clone types, five of these have previously been identified in CF patients: DK06, DK08, DK19, DK21, and DK51 [13]. This suggests that there is no single clone type(s) specifically responsible for the infections seen in either CF or PCD patients. It has previously been documented that clone types can be transmitted between CF patients [13,25], and since the Copenhagen CF clinic is located at the same hospital as the Danish PCD Centre this close proximity increases the risk of patients being exposed to the same *P. aeruginosa* clone types: This may be caused by (1) direct patient-to-patient transmission, (2) indirect transmission via environmental reservoirs or reservoirs at the hospital, or (3) clone types common in the environment [26,27].

We investigated the possibility of direct patient-to-patient transmission by measuring the genetic distance between the earliest isolates sequenced from CF [13] and PCD patients with shared clone types, (Supplementary Table 1). In all cases 68 SNPs or more separated the isolate with shared clones making a recent transmission between patients, included in the study, unlikely to have occurred, taking into account a within-patient mutation rate of *P. aeruginosa* of around 2.6 SNPs/year, as found in CF [11].

The PCD patients included in this study have previously been shown to be infected with independent clone types assessed by Pulsed-field-gel-electrophoresis (PFGE), and 10 of the 12 patients were found to be infected by a single primary clone type, whereas two patients (P01 and P06) have a possible clone type switch during the time of infection covered [5]. The whole genome sequencing of the same isolates supports these observations (Figure 1A).

Genetic adaptation

P. aeruginosa is known for its genetic and phenotypic plasticity as a microbe, as it can be found in watery environments, soil, and human airways. Despite this large repertoire, the roads of genotypic adaptation to the human airways of CF patients seem to be relatively uniform. Previously, lists of genes that have been hit by mutation as a result of positive selection during adaptive evolution, referred to here as “pathoadaptive genes” [11-13,28] have been published. Pathoadaptation is the adaptation of micro-organisms to a host environment through mutations, and includes both alterations and destructions of gene functions. In CF it appears that bacteria with multiple mutations in “pathoadaptive genes” are more resilient and more likely to persist in the host [11].

Since the mucociliary clearance is defective in both PCD and CF, many properties of the PCD and CF airway environment are similar, including the slow sputum transportation [29], the intensive antibiotic treatment, and the surrounding bacterial flora [5]. Therefore, we would expect the same genes to be hit by mutation during the course of colonisation and infection by *P. aeruginosa* in PCD and CF [11-13,28].

To identify pathoadaptive genes in *P. aeruginosa* isolated from the PCD patients, we compared the genomes of isolates from the PCD patients with the same clone type, a strategy reviewed in [30], and identified mutations that had accumulated since the earliest isolate, in this case representing the most recent common ancestor (MRCA). This was only possible for clone types that were represented in the PCD collection with more than one isolate, thus excluding clone type DK06, DK21, and DK54 from this analysis.

Pathoadaptive genes

In total we found 417 non-synonymous mutations that accumulated in the recent evolutionary history of the 11 clone types. To identify mutations relevant for the adaptation to the PCD airways we sought to find genes mutated in parallel between the clone types. These genes were identified using an extended method of that previously described [13]. The number of non-synonymous mutations accumulated in each clonal lineage was used to estimate the expected number of mutations to be found in each individual gene, depending on the size of the gene.

We identified eight genes to be mutated more often than would be expected by random distribution of mutations (Figure 2 A, Supplementary Table 2). Of these eight genes, six have previously been identified as pathoadaptive in the CF airways [11-13] (Figure 2 B). The genes found to be advantageously mutated in both CF and PCD are: *mucA*, *algU*, *lasR, mexZ, mexS,* and *mexA*. The two genes not identified in the above mentioned pathoadaptive gene lists from previous CF studies are: *pilG* and *pscP*. However, both genes have been identified as mutated in CF populations of *P. aeruginosa [13]*, and though *pilG* is not identified as pathoadaptive, other pili-processing genes such as *pilQ* and *pilD* have been recorded as pathoadaptive [13]. It should also be noted, that even though the genes *lasR* and *mucA* are not on the pathoadaptive gene list of Marvig et al. [11], they were found to have been mutated in all isolates; however, the mutations had happened decades before sampling of the first isolate. It is important to note that of the eight pathoadaptive genes three are mutations in *mex* genes. The *mex* genes encode efflux pumps that have been found to be important for the resistance towards many antibiotics [31-33]. In 20 out of the 35 isolates we find mutations in one or more of these *mex* genes. This indicates, like in CF, that a primary selection force is the antibiotic pressure.

Two of the other pathoadaptive genes have previously been shown to be constrained by historic contingency [13], which led us to investigate whether this is also the case in PCD. Six isolates were found to carry mutations in both or one of the genes, and in only one case did we find a mutation in *algU* without finding mutations in *mucA* (Supplementary Table 3), and

in this case it was a silent mutation. Otherwise four isolates had mutations in both genes, and one had a mutation in only *mucA*.

Phenotypic adaptation

To investigate whether the genotypic overlap extended to a phenotypic overlap, we tested the properties of the PCD isolates with regard to mucoidity, protease production, swimming and twitching motility, generation time, and attachment (i.e. biofilm formation) (Figure 1B, Figure 3, and Figure 4). We found that none of the longitudinal isolates showed a clear pattern of becoming mucoid, and only four out of the 12 patients seem to have populations that loose swimming motility and the ability to produce protease.

With regard to biofilm we find that the earliest isolates from each patient compared to the latest isolates remain equally capable of attachment and biofilm formation, only two patients' isolates show significant change in this trait, towards an increase in biofilm production ($p<0.01$ and $R^2>0.8$), Figure 3.

Overall, the generation time does not seem to increase and we find a median generation time of 36.33 (range: 22-104 min) similar to wild type PAO1, found in this study to have a generation time of 29.69 (1.82 Standard deviation). Four patients showed significant changes in generation time over the time of infection two increased (P05 and P12) and two decreased (P06 and P07), $p<0.05$, $R^2>0.7$.

These results are opposed to what has been shown for *P. aeruginosa* populations in CF airways, where the hallmarks of advanced CF lung infection are: Increased mucoidity, increased generation time, loss of protease production, loss of motility as well as loss of the ability to attach to polystyrene[6,9,14,15,34,35], Table 1.

DISCUSSION

We hypothesised that if the main selective pressures in CF and PCD airways were the same, we would see no difference between the genes mutated or in the phenotypic adaptive traits. The main difference between the two environments is the composition of the airway mucus, which is where the bacteria are known to be situated and proliferate. In CF the mucus has a low NaCl osmolarity and high viscosity compared to the normal mucus in the PCD airway. By genome sequencing 35 and phenotypically characterizing 41 *P. aeruginosa* isolates from PCD patients, we have been able to identify both similarities and dissimilarities in the genotypic and phenotypic adaptations between the two diseases.

We have found evidence of convergent molecular evolution in eight genes in isolates from PCD patients, where all have previously been identified as mutated in *P. aeruginosa* isolates from CF airways [13], and six have been found specifically important for the adaptation in CF airways[11-13]. Thus, the genome sequences reflect similarity of the selective pressures acting on the evolution and adaptation of *P. aeruginosa* in the airways of the two patient groups. It is also important to note the absence of specific "PCD pathoadaptive genes" in this

study, supporting the method used for the identification of pathoadaptive genes, as well as underlines the similarities of selective pressures found in the two diseases.

This similarity of the genetic adaptation, together with the difference in sputum composition indicates that the primary selective forces for the evolution of *P. aeruginosa* in the CF and PCD airways are: (1) the pressure from the immune system, (2) competition with other microbes in the airways, or (3) antibiotic treatments, since these are the common factors of the two environments.

Surprisingly, and despite the similar genomic adaptation, the phenotypic traits of the PCD isolates retained most of the *P. aeruginosa* wild type traits, such as being motile, producing proteases, maintaining and in some cases increasing the ability to attach (producing biofilm), and retain the growth rate as compared to PAO1, and what is found in CF[6].

This is consistent with a lower viscosity of the mucus of PCD patients compared to the CF patients, making it possible for *P. aeruginosa* to move. This indicates that the loss of motility, found in CF isolates over time, is not only an effect of the selective pressure from the host immune defence [8], but also a question of the selective pressures from an environment of highly viscous sputum that strongly inhibits the use of pili and flagella. The difference in viscosity of the mucus is also a plausible explanation for the increased biofilm production, where the mucus viscosity in CF might not require *P. aeruginosa* to produce its own extracellular matrix; whereas in PCD, the lower degree of viscosity of the mucus necessitates *P. aeruginosa* to produce its own extracellular matrix.

This is also reflected by the results of attachment where the PCD isolates either remain status quo or increase the biofilm formation over time, which is in opposition to the results found by Lee et al. [14] for CF isolates which show decreased biofilm formation over the time of infection.

To further support our theory, we find only a few patients with bacterial populations that loose motility. According to O'Tool [36] loss of motility is linked to a decreased ability to attach as compared to motile isolates, explaining the differences of attachment seen for PCD and CF isolates.

It could be argued that the progression of chronic infection in PCD inhibits us from detecting the adaptive change of phenotypic traits, as has been observed over time from initial colonisation to chronic infection in CF patients. If this is the case, the lack of change cannot be compared with the presence of change which has been found in CF. Still, with a median generation time of 36.33 minutes this differs significantly ($p<0.01$, T-test) from the 63.11 minutes found by Markussen et al. [6] in CF patients chronically infected with the DK2 clone type (excluding early isolates from before 1973), Figure 4.

Also, with regard to protease production, bearing in mind the long persistence of chronic infection in the included PCD patients, the isolates still produce protease. Whereas this trait has been shown to be lost in CF [6,15].

A reason for the observed differences in growth rate and protease production between CF and PCD isolates might be the ability to move within the environment, which affect these properties. Due to preserved motility the bacteria are not limited to the nutrients in close

proximity, but can “invade” new niches when necessary. In this way they increase the available nutrients and thus do not limit the growth in the same degree as in CF airways, where the bacteria probably are more stationary. This “re-infection” of new niches might also be what causes the ability to produce proteases to linger. This ability is known to be present and possibly advantageous in newly infecting isolates of *P. aeruginosa* [16] and thereby might aid the invasion of new niches of the PCD airways.

It is worth noticing, that PCD patients in many cases can clear the chronic *P. aeruginosa* infection [5]. This stands in contrast to CF, where clearance is rarely achieved. Clearance may favour re-colonisation with wild type environmental isolates and thus explain the phenotypic differences found in this study between PCD and CF isolates. However, this is contradictory to the genetic adaptations identified in this study suggesting that the phenotypic changes found in the PCD isolates are primarily caused by genotypic adaptation and mutation accumulation rather than phenotypic acclimation, indicating that sufficient time for genetic adaptation has passed.

CONCLUSION

Our results indicate that while in the CF airways the isolates quickly adapt to the lung environment by losing several phenotypic traits the isolates from PCD airways, seemingly, have an advantage if they retain their wild type-like phenotype, this despite of the similarity of genes that are mutated and identified as important for the colonisation and infection in both disease entities.

These similarities and specifically the dissimilarities between the two environments are important for the understanding of the evolution of *P. aeruginosa*. The discontinuity of genotype - phenotype, reveals a gap in the understanding of the evolution of *P. aeruginosa* to the human airways, and speaks of an evolution that in particular is characterised by a complex web of adaptational roads towards a common goal, -chronic infection of the airways. However, the convergence of genotypic adaptation found between CF and PCD reveals a common selective pressure of antibiotic treatments, which might by further investigations reveal new possibilities for treatment of *P. aeruginosa* infections.

Supplementary material:

Supplementary Table S1: Distance (number of SNPs) between initial isolates of shared clone types, from both PCD and CF patients.

Supplementary Table S2: Genes of PAO1, the number of clones observed to be mutated and expected number of clones mutated, for significantly mutated genes.

Supplementary Table S3: Historic contingencies of mutations in *mucA* and *algU.*

REFERENCES

1) Gibson RL, Burns JL, Ramsey BW. Pathophysiology and management of pulmonary infections in cystic fibrosis. *Am J Respir Crit Care Med* 2003;168:918-51.

2) Alanin MC, Johansen HK, Aanaes K, Hoiby N, Pressler T, Skov M, Nielsen KG, von BC. Simultaneous sinus and lung infections in patients with primary ciliary dyskinesia. *Acta Otolaryngol* 2014;1-6.

3) Koch C, Hoiby N. Diagnosis and treatment of cystic fibrosis. *Respiration* 2000;67:239-47.

4) Zolin A, Mckone EF, van Rens J. ECFS patient registry 2014, report 2010.

5) Alanin MC, Nielsen KG, von BC, Skov M, Aanaes K, Hoiby N, Johansen HK. A longitudinal study of lung bacterial pathogens in patients with primary ciliary dyskinesia. *Clin Microbiol Infect* 2015.

6) Markussen T, Marvig RL, Gomez-Lozano M, Aanaes K, Burleigh AE, Hoiby N, Johansen HK, Molin S, Jelsbak L. Environmental heterogeneity drives within-host diversification and evolution of Pseudomonas aeruginosa. *MBio* 2014;5:e01592-14.

7) Damkiaer S, Yang L, Molin S, Jelsbak L. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proc Natl Acad Sci U S A* 2013;110:7766-71.

8) Mahenthiralingam E, Campbell ME, Speert DP. Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis. *Infect Immun* 1994;62:596-605.

9) Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF, Tummler B. Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol* 2011;13:1690-704.

10) Nguyen D, Singh PK. Evolving stealth: genetic adaptation of Pseudomonas aeruginosa during cystic fibrosis infections. *Proc Natl Acad Sci U S A* 2006;103:8305-6.

11) Marvig RL, Johansen HK, Molin S, Jelsbak L. Genome analysis of a transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet* 2013;9:e1003741.

12) Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, Miller SI, Ramsey BW, Speert DP, Moskowitz SM, Burns JL, Kaul R, Olson MV. Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 2006;103:8487-92.

13) Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nat Genet* 2015;47:57-64.

14) Lee B, Haagensen JA, Ciofu O, Andersen JB, Hoiby N, Molin S. Heterogeneity of biofilms formed by nonmucoid Pseudomonas aeruginosa isolates from patients with cystic fibrosis. *J Clin Microbiol* 2005;43:5247-55.

15) Jiricny N, Molin S, Foster K, Diggle SP, Scanlan PD, Ghoul M, Johansen HK, Santorelli LA, Popat R, West SA, Griffin AS. Loss of social behaviours in populations of Pseudomonas aeruginosa infecting lungs of patients with cystic fibrosis. *PLoS One* 2014;9:e83124.

16) Yang L, Rau MH, Yang L, Hoiby N, Molin S, Jelsbak L. Bacterial adaptation during chronic infection revealed by independent component analysis of transcriptomic data. *BMC Microbiol* 2011;11:184.

17) Ciofu O, Lee B, Johannesson M, Hermansen NO, Meyer P, Hoiby N. Investigation of the algT operon sequence in mucoid and non-mucoid Pseudomonas aeruginosa isolates from 115 Scandinavian patients with cystic fibrosis and in 88 in vitro non-mucoid revertants. *Microbiology* 2008;154:103-13.

18) Pritt B, O'Brien L, Winn W. Mucoid Pseudomonas in cystic fibrosis. *Am J Clin Pathol* 2007;128:32-4.

19) Li Z, Kosorok MR, Farrell PM, Laxova A, West SE, Green CG, Collins J, Rock MJ, Splaingard ML. Longitudinal development of mucoid Pseudomonas aeruginosa infection and lung disease progression in children with cystic fibrosis. *JAMA* 2005;293:581-8.

20) Werner C, Onnebrink JG, Omran H. Diagnosis and management of primary ciliary dyskinesia. *Cilia* 2015;4:2.

21) Matsui H, Grubb BR, Tarran R, Randell SH, Gatzy JT, Davis CW, Boucher RC. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* 1998;95:1005-15.

22) Barbato A, Frischer T, Kuehni CE, Snijders D, Azevedo I, Baktai G, Bartoloni L, Eber E, Escribano A, Haarman E, Hesselmar B, Hogg C, Jorissen M, Lucas J, Nielsen KG, O'Callaghan C, Omran H, Pohunek P, Strippoli MP, Bush A. Primary ciliary dyskinesia: a consensus statement on diagnostic and treatment approaches in children. *Eur Respir J* 2009;34:1264-76.

23) Johansen HK, Norregaard L, Gotzsche PC, Pressler T, Koch C, Hoiby N. Antibody response to Pseudomonas aeruginosa in cystic fibrosis patients: a marker of therapeutic success?--A 30-year cohort study of survival in Danish CF patients after onset of chronic P. aeruginosa lung infection. *Pediatr Pulmonol* 2004;37:427-32.

24) Johansen HK, Moskowitz SM, Ciofu O, Pressler T, Hoiby N. Spread of colistin resistant non-mucoid Pseudomonas aeruginosa among chronically infected Danish cystic fibrosis patients. *J Cyst Fibros* 2008;7:391-7.

25) Ojeniyi B, Frederiksen B, Hoiby N. Pseudomonas aeruginosa cross-infection among patients with cystic fibrosis during a winter camp. *Pediatr Pulmonol* 2000;29:177-81.

26) Wiehlmann L, Wagner G, Cramer N, Siebert B, Gudowius P, Morales G, Kohler T, van DC, Weinel C, Slickers P, Tummler B. Population structure of Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* 2007;104:8101-6.

27) Zimakoff J, Hoiby N, Rosendal K, Guilbert JP. Epidemiology of Pseudomonas aeruginosa infection and the role of contamination of the environment in a cystic fibrosis clinic. *J Hosp Infect* 1983;4:31-40.

28) Feliziani S, Marvig RL, Lujan AM, Moyano AJ, Di Rienzo JA, Krogh JH, Molin S, Smania AM. Coexistence and within-host evolution of diversified lineages of hypermutable Pseudomonas aeruginosa in long-term cystic fibrosis infections. *PLoS Genet* 2014;10:e1004651.

29) Bush A, Payne D, Pike S, Jenkins G, Henke MO, Rubin BK. Mucus properties in children with primary ciliary dyskinesia: comparison with cystic fibrosis. *Chest* 2006;129:118-23.

30) Marvig RL, Sommer LM, Jelsbak L, Molin S, Johansen HK. Evolutionary insight from whole-genome sequencing of Pseudomonas aeruginosa from cystic fibrosis patients. *Future Microbiol* 2015;10:599-611.

31) Ziha-Zarifi I, Llanes C, Kohler T, Pechere JC, Plesiat P. In vivo emergence of multidrug-resistant mutants of Pseudomonas aeruginosa overexpressing the active efflux system MexA-MexB-OprM. *Antimicrob Agents Chemother* 1999;43:287-91.

32) Yamamoto M, Ueda A, Kudo M, Matsuo Y, Fukushima J, Nakae T, Kaneko T, Ishigatsubo Y. Role of MexZ and PA5471 in transcriptional regulation of mexXY in Pseudomonas aeruginosa. *Microbiology* 2009;155:3312-21.

33) Poole K. Multidrug efflux pumps and antimicrobial resistance in Pseudomonas aeruginosa and related organisms. *J Mol Microbiol Biotechnol* 2001;3:255-64.

34) Jorgensen KM, Wassermann T, Johansen HK, Christiansen LE, Molin S, Hoiby N, Ciofu O. Diversity of metabolic profiles of cystic fibrosis Pseudomonas aeruginosa during the early stages of lung infection. *Microbiology* 2015;161:1447-62.

35) Yang L, Jelsbak L, Marvig RL, Damkiaer S, Workman CT, Rau MH, Hansen SK, Folkesson A, Johansen HK, Ciofu O, Hoiby N, Sommer MO, Molin S. Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci U S A* 2011;108:7481-6.

36) O'Toole GA, Kolter R. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Mol Microbiol* 1998;30:295-304.

FIGURES AND TABLES

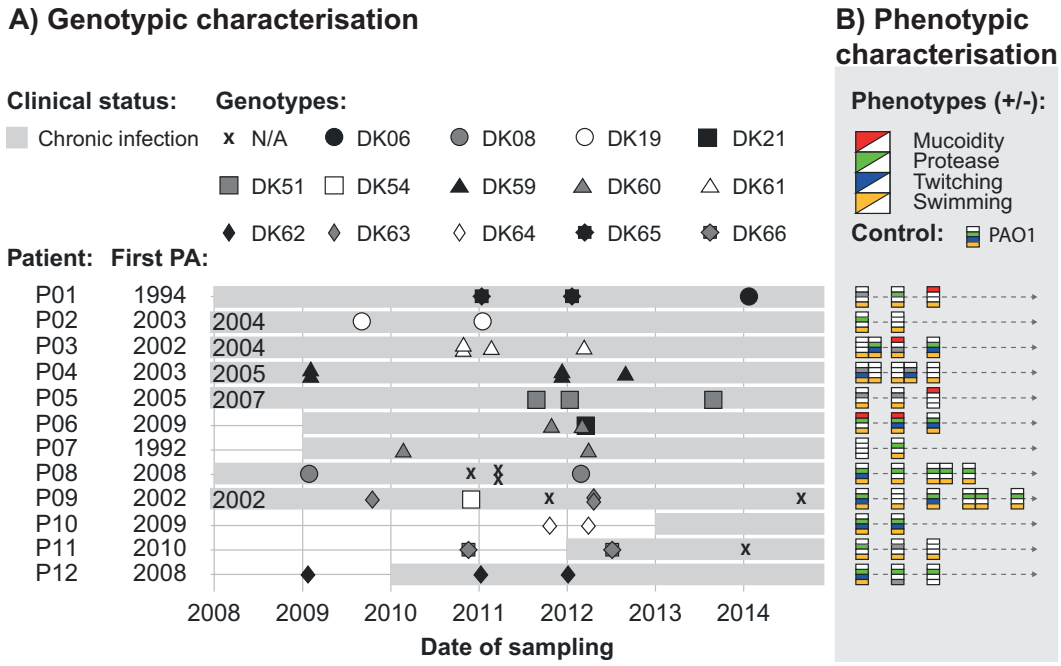
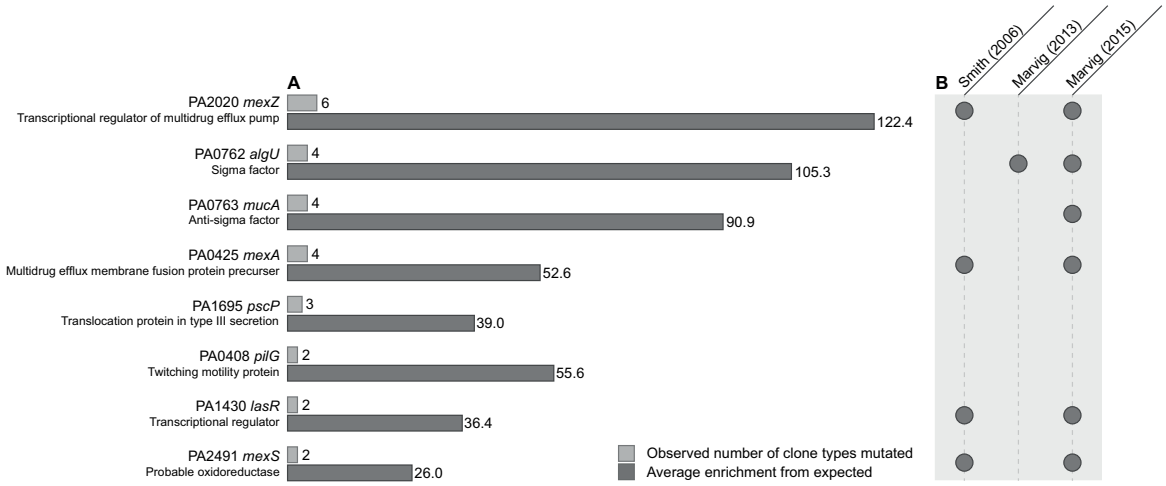
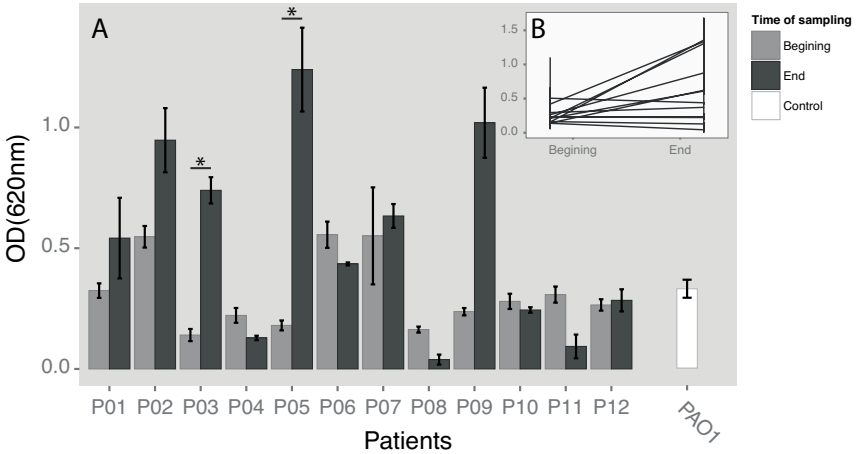


Figure 1:Patient and isolate overview. (A) Genotypic characterisation. ”First PA” shows the year the patients had the first culture of *P. aeruginosa*. When a year is noted on the grey bar (indicating chronicity of the patients) this is the year when the patient was diagnosed as chronically infected. Isolates denoted by ”X” was not sequenced. **(B)** Phenotypic characterisation. Here the phenotypic characters are shown for 41 isolates sampled at the same time points as indicated by isolates in (A). When phenotypic boxes are close together they were sampled from the same sputum sample (se sub-figure (A)). A grey box indicates an uncharacterisable phenotype (N/A), white indicates the absense of a phenotype, and coloured boxes indicate the presence of a phenotype.



500
501 **Figure 2: Genes mutated in more clones than would be expected by genetic drift (p-**
502 **values < 0.0001). (A)** Overview of the number of clones observed to be mutated in each gene,
503 and the enrichment of clones mutated in relation to expected **(B)** Overlap of the genes found
504 important for the adaptation of *P. aeruginosa* to the PCD airways, and gees found important
505 for the adaptation to CF airways in three other studies: Smith (2006), Marvig (2013), and
506 Marvig (2015).

507
508



509
510 **Figure 3: Attachement assay. (A)** Barplot of the first and last isolate from each patient
511 (Figure 1), including PAO1 as control, * $p<0.01$ and $R^2>0.8$, Linear regression with Bonferroni
512 correction for multiple testing. Errorbars: Standard error of the mean (SEM) (SD/sqrt(n), n=
513 number of samples). **(B)** Overall view of the tendency of the isolates to either remains status
514 quo with regard to biofilm production or increase.

515

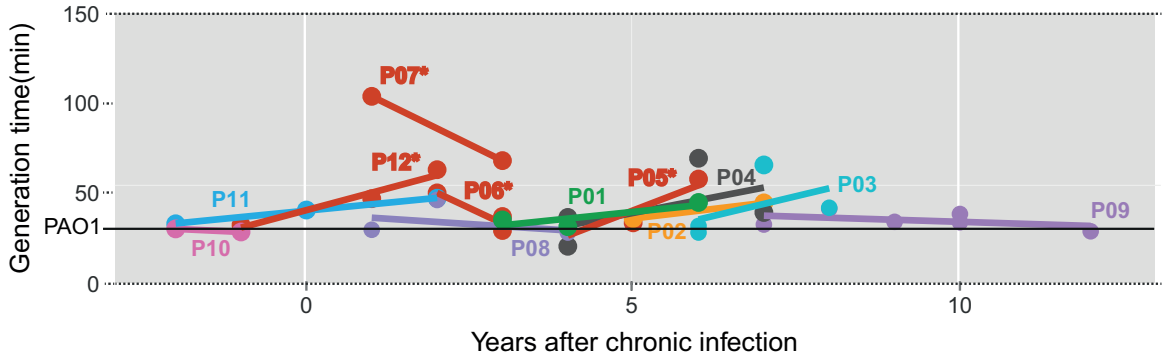


Figure 4: Generation time. Growth data from the twelve PCD patients. * and red bold patient IDs denote $p<0.05$ and $R^2>0.69$, Linear regression with Bonferroni correction for multiple testing.

Table 1: Phenotypic hallmarks of chronic lung infection with *P. aeruginosa* in patients with cystic fibrosis compared with primary ciliary dyskinesia.

Phenotypic traits	CF	PCD
Mucoidity	↗ [19], [18]	→
Generation time	↗ [6], [16]	→
Motility*	↘ [8], [6]	→
Biofilm formation	↘ [14]	↗→
Protease	↘ [15], [6]	↘ →

Study 4 Supplementary:

Supplementary Table S1

Distance (number of SNPs) between initial isolates of shared clone types, from both PCD and CF patients.

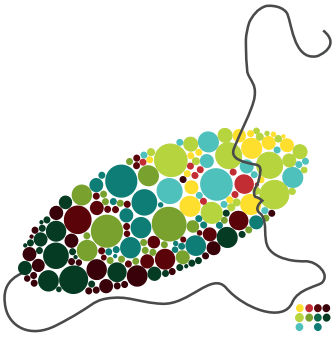
DK06	**Patient**	**P44F5**		**P21F4**		**P99F4**		**P77F4**
Patient	**Isolate ↓→**		**35**		**294**		**390**		**398**
P1	**PCD-25**		1322		102		109		102
DK08	**Patient**	**P88M4**		**P73M4**				
Patient	**Isolate ↓→**		**66**		**351**			
P08	**PCD-03**		534		220			
DK19	**Patient**	**P41M3**		**P70F4**		**P72F4**		**P55M4**
Patient	**Isolate ↓→**		**156**		**338**		**346**		**415**
P2	**PCD-05**		68		509		573		575
DK21	**Patient**	**P22M4**						
Patient	**Isolate ↓→**		**182**					
P06	**PCD-01**		107					
DK51	**Patient**	**P99F4**						
Patient	**Isolate ↓→**		**388**					
P05	**PCD-08**		165					
DK60	**Patient**	**P07**						
Patient	**Isolate ↓→**	**PCD-33**						
P06	**PCD-31**		882					
Minimum distance between shared clones (# SNPs)								
68								
Maximum distance between shared clones (# SNPs)								
1322								
Average distance between shared clones (# SNPs)								
405.2								

Supplementary Table S2
Genes of PAO1, the number of clones observed to be mutated and expected number of clones mutated, for significantly mutated genes.

Gene_ID	PA2020	PA0762	PA0763	PA0425	PA1695	PA0408	PA1430	PA2491
Gene name	mexZ	algU	mucA	mexA	pscP	pilG	lasR	mexS
Observed number of clones mutated in gene	6	4	4	4	3	2	2	2
Expected number of clones mutated in gene (minimum)	0	0	0	0	0	0	0	0
Expected number of clones mutated in gene (maximum)	1	1	2	2	2	1	1	1
Expected number of clones mutated in gene (variance)	2.35	1.41	1.89	5.70	5.85	1.26	2.97	5.85
Expected number of clones mutated in gene (average)	0.05	0.04	0.04	0.08	0.08	0.04	0.06	0.08
Average enrichment	122	105	91	53	39	56	36	26
Probality of observation if random drift [P(Z≥Y) ~ pois(Z; Y)]	1.29E-13	6.40E-10	1.32E-09	1.98E-08	1.38E-06	7.57E-06	2.66E-05	7.18E-05
Function	Transcriptional regulator of multidrug efflux pump	Sigma factor AlgU	Anti-sigma factor MucA	Multidrug efflux membrane fusion protein MexA precursor	Translocation protein in type III secretion	Twitching motility protein PilG	Transcriptional regulator LasR	Probable oxidoreductase
PseudoCAP class	Transcriptional regulators	Transcriptional regulators	Transcriptional regulators; Cell wall / LPS / capsule; Secreted Factors (toxins, enzymes, alginate)	Transport of small molecules; Antibiotic resistance and susceptibility	Protein secretion/export apparatus	Two-component regulatory systems; Chemotaxis; Motility & Attachment	Adaptation, Protection; Transcriptional regulators	Putative enzymes

Suplementary Table S3:
Historic contingencies of mutations in *mucA* and *algU*.
1 = mutation found in gene, if not insertion or deletion the mutation is written in parenthesis, mis=missense, non=nonsense, and silent. *=also contains a silent SNP.

Patient	Isolate No.	Internal ref. Number	Date of isolation	mucA	algU
P01	25	1809_11	10/01/11		
P01	14	3517_12	16/01/12		
P01	13	312_1_14	20/01/14		
P02	36	44994_09	30/08/09		
P02	33	2806_11	16/01/11		
P03	31	54878B_10	26/10/10	1(nonsense)	1(missense)
P03	**3**	54878A_10	26/10/10		
P03	5	9983_2_11	22/02/11		
P03	4	13713_12	07/03/12		
P04	29	6704B_09	01/02/09		
P04	23	6704A_09	01/02/09		
P04	41	66002B_11	07/12/11	1	1
P04	6	66002A_11	07/12/11	1	
P04	34	45477_12	20/08/12		
P05	1	2141_11	28/08/11		
P05	2	2764_12	11/01/12		
P05	7	3541_13	23/08/13		
P06	10	57224_11	26/10/11		
P06	11	12277_12	28/02/12		
P06	35	13904_12	07/03/12		
P07	40	10118_10	22/02/10		
P07	9	18506_12	28/03/12		
P08	37	3149_09	16/01/09		
P08	22	56781_10	06/11/10		
P08	18	8523_1_11	18/02/11		
P08	26	8523_2_11	18/02/11		
P08	27	10624_12	19/02/12	1(missense)	1(missense)
P09	32	53051_09	15/10/09	1*	1(missense)
P09	8	60945_10	26/11/10		
P09	15	51838_3_11	04/10/11		
P09	17	22943_3_11	18/04/12		1(silent)
P09	39	22943_1_11	18/04/12		
P09	16	3228_14	20/08/14		
P10	38	55545_11	17/10/11		
P10	30	18012_12	26/03/12		
P11	12	59511_10	18/11/10		
P11	21	37392_12	02/07/12		
P11	19	3649_14	16/09/14		
P12	28	4223_09	21/01/09		
P12	20	1037_11	06/01/11		
P12	24	3362_12	06/01/12		



Novo Nordisk Foundation Center for Biosustainability & Department of Systems Biology
Technical University of Denmark

Kogle allé 6
2970 Hørsholm
Phone: +45 45 25 80 00
http://www.biosustain.dtu.dk/